

DOCUMENT RESUME

ED 222 039

FL 013 143

AUTHOR Hellgren, Paul
TITLE Communicative Proficiency in a Foreign Language, and its Evaluation: An Analysis of the Concept and an Experiment With Oral Proficiency. Research Report 2.
INSTITUTION Helsinki Univ., (Finland). Dept. of Teacher Education.
REPORT NO ISBN-951-45-2611-2
PUB DATE 82
NOTE 182p.; Ph. D. Dissertation, University of Helsinki, Finland.
EDRS PRICE MF01/PC08 Plus Postage.
DESCRIPTORS *Communicative Competence (Languages); *Language Processing; Language Research; *Language Tests; Listening Comprehension; Psycholinguistics; Reading Skills; *Second Language Learning; Sociolinguistics; *Speech Communication; *Writing Skills

ABSTRACT

A study is presented which attempts to clarify the notion of communicative proficiency in a foreign language and which outlines an instructional model that incorporates the insights of the study. Communicative proficiency is defined as fluent mastery of the foreign language in different contexts. The definition was operationalized into discourse processing by means of an instructional model to guide teaching, choice of contents, and evaluation. An account is given of how the instructional model was used to develop a model for evaluating communicative proficiency in a foreign language. Discourse processing as a bidirectional interpreting skill is seen in terms of oral proficiency as a function of speaking and listening, and written proficiency as a function of writing and reading. A simulation of the instructional model is employed in a form in which oral proficiency is rated by written products and written proficiency by oral products. In other words, a hypothesis is made of a unitary structure of English proficiency. This hypothesis was tested by a communicative test given to third-form students in 10 senior secondary schools in Finland. The test required oral and written responses to questions about the contents of an interview heard on tape. Analysis of results showed the test to be a reliable instrument with the asset of simplicity and ease of administration. (Author/AMH)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

ED2222039

RESEARCH REPORT 2

Department of Teacher Education, University of Helsinki
Ratakatu 2, SF-00120 Helsinki 12, Finland

Paul Hellgren

COMMUNICATIVE PROFICIENCY
IN A FOREIGN LANGUAGE,
AND ITS EVALUATION

An analysis of the concept and an experiment
with oral proficiency

*Academic dissertation to be publicly discussed, by due permission of the Faculty of Education
in the University of Helsinki, in Auditorium 1 of the Department of Teacher Education,
on May 8, 1982, at 10 o'clock a.m.*

FL013 143

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

Paul Hellgren

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

X This document has been reproduced as
received from the person or organization
originating it.

Minor changes have been made to improve
reproduction quality.

• Points of view or opinions stated in this docu-
ment do not necessarily represent official NIE
position or policy.

Helsinki 1982

2

ISBN 951-45-2611-2
ISSN 0359-4203
Helsingin Yliopiston Monistuspalvelu
Painatusjaos Helsinki 1982

Contents

Acknowledgements	5
Abstract	7
Introducing the subject	9

PART I

1.	The purpose of the study	11
2.	On communication	14
3.	Analysis of meaning in the language	16
3.1.	The relationship of the linguistic symbols with reality	16
3.2.	Philosophical analysis of language in terms of communication	20
3.3.	Discourse analysis	24
3.4.	Linguistic context and situation	26
4.	Communicative competence	28
4.1.	Universal pragmatics	28
4.2.	The perspective of theoretical linguistics	31
4.3.	The perspectives of sociolinguistics and psycholinguistics	32
5.	Communicative proficiency in a foreign language	35
6.	An instructional analysis of foreign language proficiency	38
6.1.	Foreign language pedagogy	38
6.2.	A sketch on an instructional model for foreign languages	43
6.2.1.	An instructional syllabus of foreign languages	47
6.2.2.	Analysis of foreign language skills	52
6.2.2.1.	Foreign language skills in terms of communication	52
6.2.2.2.	Foreign language skills in terms of neurolinguistics	55
6.2.2.3.	Empirical analysis of language proficiency	57
6.2.3.	Proficiency in spoken language	60
6.2.3.1.	Speech perception and listening comprehension	60
6.2.3.2.	Process of speaking and speech production	63
6.2.4.	Foreign language learning in school instruction	64

PART II

7.	Evaluation of foreign language proficiency	66
7.1.	Analysis of language tests in terms of purpose	66
7.2.	Theoretical views of foreign language evaluation	67
7.2.1.	The traditionalist position	68
7.2.2.	The structuralist-behaviouristic position	68

7.2.3.	The psycholinguistic and sociolinguistic position	72
7.2.4.	Functional model for evaluation of foreign language proficiency	74
7.2.5.	Quantification of language proficiency	77
7.2.6.	Empirical studies of oral proficiency	81
7.2.6.1.	Previous listening comprehension tests	81
7.2.6.2.	Previous speaking tests	85

PART III

8.	The experiment	92
8.1.	Defining the purpose of the experimental study	92
8.2.	Problems and hypotheses of the experiment	94
8.3.	Preparing the measuring instruments	96
8.3.1.	The communicative test	96
8.3.2.	The cloze test	98
8.3.3.	The listening comprehension test	100
8.3.4.	The questionnaire	101
8.4.	The population of the experimental investigation	102
8.4.1.	Sampling technique	104
8.4.2.	Sampling of schools	106
8.4.3.	Sampling of students	108
8.5.	Data gathering	108
8.6.	Scoring the instruments	110
8.6.1.	Scoring the communicative test	110
8.6.2.	Scoring the cloze test	113
8.6.3.	Scoring the questionnaire	113
8.7.	Results	114
8.7.1.	The first problem area	114
8.7.2.	The second problem area	119
8.7.3.	The third problem area	129
8.8.	Discussion of the experiment	139
9.	Conclusion of the research	148
	References	151
	Appendices	159

Acknowledgements

My sincerest thanks and acknowledgements are to be extended to all those who contributed to this research with their professional expertise and constructive criticism, support and encouragement, exposure to the experiment and practical aid in accomplishing the research and in completing this report. Among them I should like to mention the following names in particular.

Professor Erkki A. Niskanen supported me by his encouragement and by creating favourable working circumstances. Professor Pertti Kansanen generously discussed various research problems. He was also one of the efficient readers of the manuscript of this research. The other was Associate Professor Anna-Liisa Leino, who, as the expert of foreign language pedagogy, followed the progress of the research and inspired me from the very beginning. Discussions about certain research problems and her official report on an earlier version of the study all contributed positively to my working. The other official report was given by Associate Professor Sirppa Kauppinen, who also stimulated my research by her rating. Dr. Toivo Heikkurinen gave his critical comments on the draft of the earlier version and thus helped to improve the development of thoughts and presentation. Dr. Juhani Jussila, as acting Professor at the Department of Teacher Education, provided me with the tapes for the experiment, created good working conditions, and encouraged my study with his reinforcement. He also read and commented on the manuscript. Associate Professor Erkki Komulainen supported me by helping to solve some statistical problems. For the preparation of the testing material, I received aid in different forms. Mr Peter Bull, M.A., acted as the interviewee and recorded the original interview with his wife. Mr Juhani Kyllönen, M.A., edited and refined the final recording of the testing material. Ms Anneli Heinonen, M.A., provided me with the file of the senior secondary schools equipped with language laboratories. The Board of the Matriculation Examination gave the data for the results from the examination of Spring 1981. In the ten senior secondary schools in which the experiment was carried out the English teachers generously cooperated and, together with their students, contributed to the accomplishment of the experiment. In preparing this report, I was helped by Ms Ritva Järvinen, B.A., who drew the figures, and by Mrs Irene Kristiansen, M.A. and Mrs Kaarina Yli-Renko, M.A., who proof-read the text.

To all the above persons I acknowledge my debt, and express my gratitude in the form of this research report. Also, I am very grateful that this study will be included in the new series of Research Reports of the Department of Teacher Education, for which I am indebted to Professor Pertti Kansanen.

Paul Helligren

Abstract

Paul Helligren

Communicative proficiency in a foreign language, and its evaluation. An analysis of the concept and an experiment with oral proficiency. Research Report 2. The Department of Teacher Education, University of Helsinki. 1982. 182 p. ISBN 951-45-2611-2. ISSN 0359-4203.

Communicative proficiency in a foreign language as a learning goal was analyzed from the perspective of foreign language pedagogy, and its achievement was evaluated in the form of oral proficiency. The study consists of three parts:

I. The concept of communicative proficiency in a foreign language was analyzed. As a result, it was defined as fluent mastery of the foreign language in different contexts, which then was operationalized into discourse processing by means of an instructional model to guide teaching, choice of contents, and evaluation.

II. The instructional model was utilized to develop a model for evaluating communicative proficiency in a foreign language. Discourse processing as a bidirectional interpreting skill manifests itself in the communicative proficiency as two functions: oral proficiency as a function of speaking and listening, and written proficiency as a function of writing and reading. In practice, a simulation of the model can be employed in the form in which oral proficiency is rated by written products, and written proficiency by oral products.

III. In the field experiment the hypothesis of the unitary structure of English proficiency was tested by means of a communicative test, in which oral and written responses were given, in a restricted time, to questions about the contents of the interview heard on tape. The test consisted of two parts with ten questions in each. The subjects ($N=406$) were third-form students of ten senior secondary schools in Finland, drawn through stratified sampling. The subjects were matched by means of a cloze test. The matched pairs answered the same questions in the different modes and changed their answering modes in the second part of the test. Differences between written and oral responses were investigated through an analysis of variance and through factoring into a two-factor solution. The hypothesis could not be rejected, and the second factor explained only 3.3—3.6 percent of the total variance of the variables. The communicative test proved a reliable and valid instrument with an asset of simplicity and administrability. The advantage of the written version comes up when, for example, a large number of testees have to be examined in oral proficiency.

A questionnaire was also given to the students to control some

background variables and to explore some tentative connexions between English proficiency and some factors within the framework of the instructional model. Among the most interesting findings can be cited: no difference in performance was found between boys and girls whereas those who had chosen the long course of mathematics performed slightly better. Students' input in out-of-class activities had a statistically significant relation with performance but their occurrence among students was not very frequent. Attitudes to the teacher and to instruction were markedly positive, but no connexion with performance was found. The use of native/foreign language in instruction did not correlate significantly with performance, either.

Introducing the Subject

The term 'communication' with its various derivatives has become almost a slogan in recent years. Verbal communication may mean a person's ability to participate in a conversation, that is, to express what he wants to say in order to be understood, and to understand what the others say. In ordinary language, communication is often associated with spoken language only. On the other hand, grammaticality may be underrated at the cost of the 'communicativeness', if the understanding of the message is the adequate criterium and the possible language mistakes are omitted.

Communicative skills have been discussed in terms of the native and foreign languages. The educational aims of the native language are often seen to broadly include a person's self-expression in various forms. Communicative proficiency in a foreign language as the educational aim has led to an emphasis on spoken language — maybe as a reaction to the traditionally stressed aim of correct written proficiency. These swings of the pendulum may only be illusionary, for curriculum designers and planners may have had a broad interpretation of communicative proficiency in their mind including all verbal skills in the concept.

Linguistic research has defined such concepts as linguistic competence and performance as well as communicative competence. These theoretical concepts of linguistics have then won a wider use in literature. They have been applied to the nomenclature of teaching and learning without much further analysis. The concepts defined in various fields have been utilized as if they were commensurable, resulting in some confusion. Talking of communicative skills might suggest that, as an educational aim, they are the same for the native and foreign languages. Generally speaking, it may not be realistic to think that a person will master the foreign language repertoire with the same sovereignty as her native language.

In the present study an attempt will be made to find order in this jungle of concepts. As a result we hope we may understand the learning goal 'communicative proficiency in a foreign language' better in terms of education, which then will facilitate organizing instruction in a meaningful way to achieve it.

The following discussion will use the terms 'first language', 'native language' and 'mother tongue' synonymously, so that they refer to the language that the person has first acquired. There may be borderline cases in which the person has acquired two languages simultaneously in her home. In our nomenclature this is bilingualism but it will not be dealt with in the present study. The term 'foreign language' runs through the study referring to the language that a person has learnt or acquired or is learning or acquiring after the first language. The term has the same

meaning as the frequently used 'second language'. The latter term does not seem quite appropriate because it suggests only another language, in addition to the first one. In Finland, it is common to study more than one 'second language', which makes it more convenient to stick to 'foreign language'.

The term 'foreign language' has no connotation of the level of performance or of the way in which it is learnt or acquired. Neither does it refer to its position in a country. The foreign language is seen from the perspective of the individual. We find this view relevant to foreign language instruction. In Finland Finnish and Swedish are the official languages of the country. Swedish is, however, a foreign language to those who must learn it at school. We may of course speak of Swedish or Finnish as the second language of the country. The advantage of this distinction is that Swedish, for example, can be included in the foreign language pedagogy when it is not taught as the student's first language at school. Some writers refer with the term 'bilingual' to the person who has acquired the fluent mastery of a foreign language. We do not find it consistent with our earlier definition of bilingualism, so that it will not be used in this study.

PART I

1. The Purpose of the Study

This study intends to elucidate communicative proficiency in a foreign language. The research lies within foreign language pedagogy. It follows that the analysis of communicative proficiency will be aimed at outlining an instructional model for foreign language teaching. With the emphasis on instruction, it is hoped that the teaching of foreign languages in schools will benefit from the analysis, because communicative proficiency is studied from the perspective of teaching, and not as an application of linguistic or psychological theories. Foreign language pedagogy as an interdisciplinary field of study may easily be carried away into a jungle of concepts, which makes it difficult to shape meaningful models for action. On the other hand, communicative proficiency must be placed in a framework which provides the basis for evaluating it as an outcome of learning.

The problems of this research on proficiency in a foreign language are, in detail, as follows:

1. What does communicative proficiency in a foreign language mean conceptually as a learning aim?
2. How can communicative proficiency in a foreign language be put into instructional practice, i.e. what instructional model might efficiently direct teaching and research on it?
3. How can communicative proficiency be evaluated in terms of the instructional model?
4. How can communicative oral proficiency be measured in terms of the model for evaluation?

The first question will be answered by means of a conceptual analysis. Language will be analyzed from the viewpoint of communication, and an attempt is made to distinguish the constituent parts that a foreign language learner must acquire in order to be able to communicate in the language. This serves as a basis for the definition of communicative proficiency as a learning aim. The outcome of the analysis will be utilized to outline an instructional model, which sketches a framework for foreign language teaching in schools and for its investigation.

The third question will be answered by analyzing the separate skills of a foreign language in terms of the instructional model. The result will be used to define the frame of reference for the empirical assessment of communicative proficiency in a foreign language. The theoretical frame will help to develop a model for evaluation, which then guides the construction of measuring instruments.

The fourth question will be answered by means of an experiment. A

new type of oral test that can be given even to a large number of subjects will be tried out and investigated. Comprehension of spoken English is rated by means of written response to questions about the contents of the spoken text heard on the tape. This is considered to simulate oral response except that pronunciation and intonation are omitted. It is supposed that in this form the test type is capable of measuring oral language proficiency.

The experiment will be confined to proficiency of English achieved after several years of studies, by the end of the senior secondary school. This stage can be justified by the fact that the level is the most advanced in terms of the curricular aims for schools. The school-leaving examination, also called the Matriculation Examination, is based upon it and it is the same for all secondary schools in Finland. The annual examination is arranged simultaneously in every school during February — early April with a renewal in September. At this phase the level of proficiency in English will remain the highest possible ever achieved for some students, offer a vehicle for further studies or serve as a basis for advancement in English. Among other things, this research will preclude learning theories and teaching techniques related to communicative proficiency.

The theoretical problems above will be approached by conceptual analysis. Concepts and theories can describe reality, combine research outcomes, direct research and practice. The existent concept of communicative proficiency in a foreign language, which frequently has been called communicative competence, will be analyzed for the purpose of being able to better understand and interpret reality.

It is possible to distinguish the intension and extension of a concept. The intension is determined by the characteristics present in the definition of the term, and by it, extension is meant all those phenomena that have these characteristics. For this study, the intension means investigations into the relation between communication and language, and into the communicative meaning manifest in the linguistic code. An attempt is made to isolate the features that are revealed in the linguistic code and that are to be inferred from it. Thus we may determine characteristics that are relevant to learning a foreign language. A useful concept is theoretical as well as practical. When communicative proficiency in a foreign language has also been defined instructionally, i.e. made operational, the concept can be considered to meet the criteria of utility.

The experimental part of the study rests upon the foregoing theoretical development but forms a whole of its own. This is because the experiment will partly have its own theoretical frame, due to the concentrated focus and to a different approach. Figure 1 depicts the scheme for the progress of this research, and the interrelations of the different stages towards the focus.

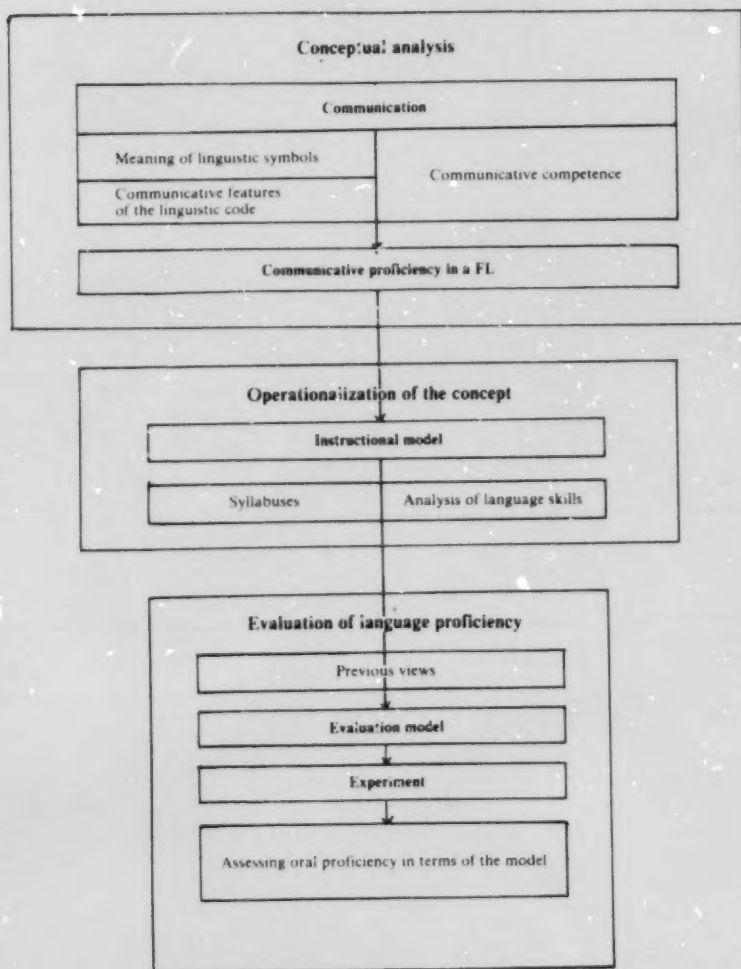


Figure 1. The scheme for the progress of this research.

Roughly speaking, the present study combines two research approaches: hermeneutic and positivistic. By hermeneutics is meant an analysis and interpretation of a phenomenon so that it adds to understanding the whole. Understanding the details and comprehending the whole proceed in support of each other, thus forming a hermeneutic circle (cf. Malinen 1974a, 14). Features characteristic of the positivistic approach can be said to be as follows: (i) logic atomism, i.e. the object is dispersed into variables which are logically independent of each other, (ii) measurability, (iii) experimentalism, and (iv) mathematical concept formation. In this study, the conceptual analysis is based upon the hermeneutic approach and the experiment on the positivistic one.

There are certain logical and practical difficulties in the attempt to solve a problem through different approaches in the same research. The ultimate aim is to tackle a practical problem, but to do so with justification requires the development of a practicable theory. Despite the difficulties and the difference of the approaches, they were considered necessary for gaining insight into our research problems.

2. On Communication

There are and have been many definitions of human communication. The definitions have changed somewhat over the years, but the changes are due more to the perspective of the scholar than to basic differences in the conceptualization of the human communication process. Although communication scholars have talked about communication as a process, the process has been used in different senses. It has been seen in the form that everything is relative and everything is related to everything else in noncodable ways. The process is then viewed as a mystery — something dynamic but undefinable. The process has also been defined as complex organization, as change over time, or as an activity — processing. Here we adopt the view developed by Berlo (1977) that communication as a process is an activity. It is an approach that incorporates some of the merits in other interpretations.

Communication activity is rooted in natural events, i.e. all symbols for which users have significance are embedded in units of matter and energy which constitute the sensory messages we exchange. Though message events are natural events, the communicative significance is not natural but symbolic. (Berlo 1977).

A symbolic relationship is social. What makes communication social is that it occurs through social coding and is subject to a history and future of shared agreement as to the coding rules. Consequently it does not refer to the requirement of two people or the awareness of one

another. Also, an information-communication relationship may be directional but it need not be. A relationship is directional if we look on the "source" as intentional and initiatory and the "receiver" as passive. But if both users approach the engagement with expectations, plans, and anticipations, the uncertainty reduction attributable to the contact may be better understood in terms of how both use and approach a message-event than in terms of how one person uses the contact to direct the other. (Berlo 1977).

Berlo's definition of 'social' helps us to rise above the actual communication situation so far that the traditionally viewed relationship between the sender and the receiver in time sequence as well as the cause and effect relationship between them may be eliminated from the definition. What both directional and undirectional communication share is its intentionality. It may be claimed that a person's expectations, plans, and anticipations are part of his intentionality. With an emphasis on intentionality, it is possible to exclude unintentional signals from the definition of communication.

There are various ways available for man to communicate, the most important of which is language. Gestures, facial expressions, clothes, etc. are means of conveying symbols, but whether they are intentional or not is not unequivocal. We can view them from another standpoint to rule them out from the definition of verbal communication.

Nonverbal communicating activities which accompany verbal interaction in conversation are called paralinguistic phenomena. Paralinguistic activities can be confined to those that communicate and that are part of conversational interaction. Thus, the two requirements rule out, for example, a nervous twitch of the eyelid and other mannerisms and tics.

Abercrombie (1973) divides paralinguistic elements into visible movements and postures, and audible movements and postures. Another dichotomy can be found in terms of function: paralinguistic elements can be independent of the verbal elements of conversation or they must be dependent on them. For example, a participant may nod his head at the same time as he says 'yes' or he may not without saying anything and thus communicate. Much of dependent visible paralinguistic behaviour comes under the heading of posture, also gesture, facial expression, and eye contacts between the participants make their contribution to the interaction. (Abercrombie 1973).

The dichotomy of independent and dependent applies to audible paralinguistic phenomena, too. Independent audible elements are interjections, which are characterized by the fact that they do not follow the normal phonological rules of the language. Tones of voice are an example of dependent audible elements, and they are extremely varied in character. (Abercrombie 1973).

Paralinguistic phenomena are neither idiosyncratic and personal nor

generally human, but culturally determined so that they differ from social group to social group. The meaning of the language is not bound by paralinguistic phenomena. Neither is it by other bodily movements under the heading of kinesics. Consequently, it is possible to omit such elements from the definition of verbal communication.

3. Analysis of Meaning in the Language

Language constitutes the symbols for verbal communication. It has two important functions. First, it is man's most essential means of interpersonal communication. Second, by means of language man structures the sense-data of the world. The latter function may be called man's internal communication (Koort 1974, 58). Thinking and language have an intimate connexion. Though there does not exist an agreement about this relationship, we may claim that the ability to speak and understand language as well as the ability to think are somehow interdependent.

Symbols stand for meanings. There are many theories of meaning, which are formulated through determining a position on the significance and interrelations of sense-data, words and sentences, the use of words and sentences in ordinary language, ideas, and the problem of reference in regard to words, sentences and ideas. Semantics is a branch of study that is concerned with meaning. Current linguistic research of meaning is divided as to whether the concern is the literal sense of the sentence or the illocutionary force of utterance. Each derives insight from distinct semantic theories in philosophy.

Symbols differ from language to language, but what is more important, connotations and denotations of words vary to such an extent that meanings in different languages do not correspond to each other. For communication in foreign languages, it is necessary to know such semantic differences as well as differences in reference rules. Therefore, the following discussion will deal with the relationship between symbols and reality, and the linguistic meaning in terms of communication.

3.1 The relationship of the linguistic symbols with reality

When we acquire a language, we also learn to participate in a social system that defines roles, manners, morals and institutions. This has made some researchers claim that a language locks us into a way of viewing the world. That manifests itself in the very categories embedded in the language itself. Such thoughts were presented by the American anthropologist B.L. Whorf, who claimed that differences in the surface

structure of a language reflect differences in semantic and conceptual levels. According to his theory of linguistic relativity, different languages determine their users' different world views and the way in which they think. He appeared to believe that the content of thought influences the process of thought so that different contents actually produce differing species of processes. Whorf based his hypothesis upon his research among the Hopi Indians.

Carroll (1963) argues against Whorf's hypothesis with rebutting evidence from language acquisition. As a result of his everyday interaction with the environment, the child has acquired a number of perceptual invariants, i.e. the capacity to recognize certain things, qualities, or events as being in some sense identical or 'invariant' on different occasions or in different presentations. In the process of acquiring language the child learns to attach labels to the perceptual invariants he has already acquired. In many cases the labels will correspond almost exactly to those perceptual invariants. Therefore Carroll (1963a) modifies the hypothesis of linguistic relativity as follows: "Insofar as languages differ in the ways they encode objective experiences, language users tend to sort out and distinguish experiences differently according to the categories provided by their respective languages. These cognitions will tend to have certain effects on behaviour."

Universalists tend to see similarities between different languages. Among them Chomsky (1965) distinguishes the surface structure of the language from the underlying deep structure. His theory suggests that despite differences in the surface structure, similarities prevail in the deep structure. Since the semantic component is assumed to operate in the deep structure, rather than in the surface structure, semantic universality between languages prevails in the deep structure. This view is thus in discord with Whorf's hypothesis.

Among the linguists, F. de Saussure's concept of linguistic sign reveals implications similar to Whorf's hypothesis. According to him, the sign was a combination of a concept and an acoustic image. The concept is the content (signified) of the sign; the acoustic image is the form (expression, signifier) of the sign, which also can be a graphic image. The two are so intimately connected that one seems inconceivable without the other. (Malmberg 1966, 10). The whole of the sign refers to the referent. Though the relationship between the sign and the referent is conventional, the sign is arbitrary in regard to the form and content on one hand, and to the form and referent on the other hand.

According to this linguistic theory, languages dissect 'reality' in different ways; for example, time, colours, family relations are expressed differently in different languages. Even grammatical classifications and systems differ. Each language is unique in its' structure and the possibility of language universals is precluded.

If de Saussure's thesis of the arbitrary linguistic sign and Whorf's hypothesis are taken literally, it will lead to the conclusion that the content structures vary from language to language. As a consequence, the systems of concepts are different, and the symbols do not correspond to each other. Nevertheless, Whorf's bold proposition runs the risk of being overstated and oversimplified. Admittedly, there is an intimate interaction between language, culture, and personality development, but it must be more complex than Whorf's hypothesis would predict. Also, the implications of de Saussure's sign need not be taken literally. Many different languages share structures and scopes of meanings. Tight historical, cultural and social relations have created a broad common basis for mutual agreement about the meanings of concepts, e.g. in Europe. Besides, our basic linguistic categories do not necessarily negate our ability to think along certain lines nor do they make it impossible for an individual to create new categories.

To exemplify some implications, let us take an example from the Finnish language. The pronoun *hän* ('he, she') stands for both sexes, unlike the corresponding word in many other languages. It does not follow that the Finnish speaking people would not be able to see the difference. It is only that speakers of some language do not pay attention to the differences that speakers of other languages make. It is not necessarily always so, for the differences can be detected and talked about in any language, though they are not always manifest in experiences. In the same way, learning new things, analyzing and forming concepts, making discoveries, etc., would be much more difficult if thinking and mental processes were completely restricted by the language.

A view more interesting than those given by linguists is provided by a philosophical description of language and its relation with reality. In his later philosophy Wittgenstein reflects upon the problem. The individual words name objects and things. Combinations of words make sentences. In reality language is not this simple, but in abstracting from these examples real language is surrounded by a haze. Wittgenstein speaks of a primitive language as a language-game (*Sprachspiel*). Such language-games can include meanings defined ostensively, i.e. they are understood by pointing to a table, a chair, etc. Meaning can also be learnt by means of a pattern (colours) or by use ('there'), which first is demonstrated ostensively. Thus there are many ways in which the word can mean something. Also, the whole, consisting of language and the actions into which it is woven, is called the language-game. In different language-games words obtain their meanings. Therefore, Wittgenstein (1953) regards that for a large class of cases the word 'meaning' can be defined thus: the meaning of a word is its use in the language (§ 43).

Words have different levels of abstractness. At the highest level meaning can only be demonstrated in connexion with other symbols of

the language. This relatedness manifests itself in the relationships among the language-games. According to Wittgenstein (1953, § 23), there are countless kinds of language-games. There are not only assertions, questions, commands, but "countless different kinds of use of what we call 'symbols', 'words', 'sentences'. And this multiplicity is not something fixed, given once for all; but new types of language, new language-games, as we may say, come into existence, and others become obsolete and get forgotten." The following exemplify language-games: "Giving orders, and obeying them — Describing the appearance of an object, or giving its measurements — ... Reporting an event — ... Forming and testing a hypothesis — ... Making up a story; and reading it — Play-acting — ... Translating from one language into another — Asking, thanking, cursing, greeting, praying." (Wittgenstein 1953, § 23).

Differences of levels in the language-games leave unsolved and vague what part situational factors and the actual language code play in the language-games. Is it implied that the situation determines the meaning of the word? This may be so in some cases, but many situations are general (e.g. describing the appearance of an object or reading a story, etc.), which makes it unjustifiable to suppose that the meaning of the word would be determined by the situation. Instead, these games obviously shape the style and the way in which the language is rendered.

About the language-games, Wittgenstein says that "I am saying, that these phenomena have no one thing in common which makes us use the same word for all, — but that they are *related* to one another in many different ways. And it is because of this relationship, or these relationships, that we call them all 'language'." (Wittgenstein 1953, § 65). A network of similarities overlapping and crisscrossing are called 'family resemblances' (Familienähnlichkeiten) because no similarity characterizes all members of a family. Wittgenstein compares his concept of family to a thread which has its fibres intertwined. The strength of the thread does not reside in some one fibre that runs through its whole length, but in the overlapping fibres.

The use of the word in the language-game shapes the meaning and the concept that people agree about. This agreement is not about opinions but about the form of life. A person learnt the concept 'pain' when he acquired the language. If language is used for communication, there must be agreement in both definitions and judgements. Though Wittgenstein thinks that language constructs the frame of man's consciousness, it would not obviously be interpreted along Whorf's way of thinking. For he says that "Thought is surrounded by a halo. — Its essence, logic, presents an order, in fact the a priori order of the world: that is, the order of *possibilities*, which must be common to both world and thought. But this order, it seems, must be *utterly simple*. It is *prior* to all experience, must run through all experience; no empirical cloudiness or uncertainty can be allowed to affect it — It must rather be of the

purest crystal." (Wittgenstein 1953, § 97). In the first place, this view can be compared to the selection of perceptual invariants and the universal features of different languages.

A comparison of language with games brings into prominence the rules that define the game and that the players obey. They are learnt and they manifest themselves when the game is being played. Language is acquired by participating in a great number of various language-games and not by first learning certain prescriptive rules which control language use on various occasions.

Wittgenstein (1953) distinguishes two aspects of the rules, but he does not define them as Searle (1969), who speaks of regulative rules and constitutive rules. The former refer to the conditions of occurrence of certain behaviour patterns, "Children must not play football on the grass." The constitutive rules constitute an activity the existence of which is logically dependent on the rules, "A player is offside if...". If children break the former rule, they are still playing football (as lawbreakers), whereas by breaking the offside rule they are not technically playing football any longer, because there is no game of football without the constitutive rules.

In the language-games we follow the constitutive rules which are obeyed blindly. The relationship between language and reality appears in the fact that when we think in the mother tongue, 'meanings' do not go through our mind in addition to the verbal expressions, but the language is itself the vehicle of thought. In thoughts the words occur in the order in which they are thought, and the order is that of the word order of the language. (Wittgenstein 1953, §§ 329, 336). Against this we could claim that the word order is hardly that of a speaker's language in thoughts but that in uttering the thought it takes the word order of the speaker's language.

Since the language-games vary a great deal in terms of levels, both types of rules are needed when the games are being played. Regulative rules are obviously followed according to the various situations, and they can be changed consciously. Changes in the constitutive rules of the language may be thought to take place as a result of a gradual unconscious development.

3.2. Philosophical analysis of language in terms of communication

The traditional grammar divided the sentences into different types according to their form, for example: indicative, imperative, interrogative. Such a distinction is insufficient to describe the functions of the sentences. The rules of grammar are simplified formal abstractions for structuring language. However, the regularities thus observed hardly apply to all communication as such or cover all the functions manifest in communication.

Austin (1962) attempted to investigate how a linguistic expression applies to each situation. After many years of observations and investigations he arrived at the synthesis that is called the theory of speech acts. Both Wittgenstein's language-games and Austin's theory of speech acts relate the functions of the language to the social contacts in which languages operate.

In developing his theory, Austin (1962) wanted first to distinguish descriptive and performative expressions. The former are sentences of statement which describe an event, process or state of affairs. For example, saying 'I eat' does not mean that I am eating. However, by using a performative verb I am also performing an act, e.g. 'I baptize this child Elizabeth', 'I promise to give up smoking.' Austin distinguished two kinds of performative expressions, implicit and explicit. For example, 'I'll be there at two o'clock' and 'I promise to be there at two o'clock' (Austin 1962, 69). The first example is implicitly and the second explicitly performative, because only the latter includes the performative verb 'promise'.

Austin (1962) divides the speech acts into locutionary, illocutionary and perlocutionary acts. A locution is performed when an utterance with a meaning is given out. Such an act can be viewed from different perspectives. Austin's definition does not tenably classify locutions from illocutions. An illocutionary act is done at the same time something is said, for example, one states, promises, commands or requests, asks, baptizes, describes. In saying 'Shut the window', I perform a locution and simultaneously I perform an illocution by request.

In a way, speech acts are a generalization of performative usage when all utterances, including statements, are regarded as performatives. Though the originally given reason for the concept of 'performative' thus disappears, the distinction between explicit and implicit performatives remains (Lyons 1977 II, 743). It is not, however, certain on what grounds Austin made this distinction. Is it necessary, for example, to use a performative verb for an explicit performative? In principle, it need not be so, though Austin's thoughts give the impression.

By locution is brought about a perlocutionary act. 'Shut the window' possibly makes the person whom I address shut the window. The illocutionary meaning of an utterance is incorporated in the very act as a promise, threat, exhortation, whereas the purpose of the perlocution is to influence the hearer's action, thoughts, beliefs, etc. For example, if I ask someone to stay for dinner, I perform an illocution, but if I persuade him to stay for dinner, I perform a perlocution. Of a perlocutionary act, it is always possible to ask whether it succeeded but not of an illocutionary act. This distinction is rather unclear because what is brought about can be incorporated in the sense of the verb but can also be a result that is not meant by the speaker or that is brought about by

means of any sentence. 'There's a bull behind you' can be a statement or a warning, or it can make someone frightened. For this study, the perlocutions are not essential.

Searle (1969) has continued to develop the theory of speech acts. He starts from the hypothesis that language is intentional behaviour determined by rules. Thus his as well as Austin's language theory is part of the theory of action. The speech acts are analyzed from the perspective of the constitutive rules, which means that intentionality and its linguistic realization are examined.

Searle (1969) distinguishes three kinds of speech acts: (a) utterance acts: uttering words (morphemes, sentences); (b) propositional acts: referring and predicating; (c) illocutionary acts: stating, questioning, commanding, promising, etc. The following are examples:

1. Sam smokes habitually.
2. Does Sam smoke habitually?
3. Sam, smoke habitually!
4. Would that Sam smoked habitually.

Each example constitutes an English sentence as uttered. All the utterances have in common that the speaker refers to a certain object, Sam, of whom he says something with the expression 'smokes habitually'. The same reference and predication occur as part of the complete speech act, which, however, is different in each example. In the first example the speaker is making an assertion, in the second asking a question, in the third giving an order, and in the fourth (an archaic form) expressing a wish. All these acts are called illocutions. Illocutionary acts are expressed by such verbs as state, describe, claim, warn, remark, order, command, apologize, promise, demand. (Searle 1969, 22—24).

The proposition is a problematic concept. Searle's proposition comprises reference and predication. Frege means by it approximately the same as the idea ('Sinn') of the sentence (von Wright 1975, 192-4), but a proposition can also mean the content of the sentence (Ibid. p. 134-5). To avoid confusion this report follows Searle's definition of the proposition, and by the content we mean a compound whole of both the propositional and illocutionary acts.

Both the constitutive and regulative rules are important in use of the language. All interaction operates by means of the regulative rules, even though they are not explicitly expressed in greetings, choice of topics, interruption, etc. The constitutive rules of the language govern the ways in which a certain form of the sentence is understood as an illocutionary act. Searle (1969) starts from the hypothesis that the semantics of the language can be regarded as a series of systems of constitutive rules and the illocutions are performed according to these series of the constitutive rules. There has been some criticism against this view. Stampe (1975) elaborates reasons that the identification of force-determining with meaning-determining rules cannot be sustained. He argues that the

identification cannot be sustained because it yields an utterly incoherent notion of a semantic rule.

Cooper (1973) argues that Wittgenstein's slogan "Meaning is use" embraces three merits. An analysis of concepts becomes an analysis of meaning, which implicitly denies the internal meaning to be sought by means of mentalistic processes. It also denies that the meanings are abstract entities expressed by words. The theory of speech acts incorporated in the theory of use enables an analysis of the meaning of utterances. Likewise Searle (1969, 18) regards every speech as determined by a given sentence or set of sentences. For this reason, a study of the meaning of sentences is not distinct from a study of speech acts.

Thus there exists a natural relationship between what the speaker means by his utterance and what his utterance means, because the latter is to a great extent explained by means of the former. This is naturally confined to what can be expressed in the linguistic code. The speaker's other intentions are apparently dependent on the situation, and other external factors. Such effects are e.g. perlocutionary in character.

The question of whether convention is necessary for the illocutionary meaning of an utterance is also controversial. Strawson (1964), following Grice, claims that the basic illocutions, such as stating, questioning, commanding, are essentially nonconventional in the sense that they can be explained solely as so-called natural responses. Searle maintains that at least some acts, e.g. asserting and promising, can only be performed by the constitutive rules of the language. The linguistic conventions are conventional realizations of these rules. (Searle 1969, 38—40). Stampe (1975) argues against this view. He wants to reject the entire doctrine that illocutionary force is determined by convention. He cannot accept the view that an illocutionary act is constituted an act of its specific kind by certain 'constitutive' rules.

It is somewhat difficult to argue for or against these different views without going deeper into the problem. For the purpose of this research, we may say that the distinction of the two kinds of rules provide a clear-cut perspective in regard to the language, even though the rules may not account for the meaning of the sentence completely. The constitutive rules may be said to operate nearer the surface structure of the language and consequently be considered conventional. Promising, asserting, asking questions, etc. are apparently functions that operate at the deeper level, which makes them nonconventional.

On the whole, Lyons states of Austin's theory that it offers a unified theory of the meaning of utterances with the framework of a general theory of social activity. His theory of meaning, like Wittgenstein's, can be described as a contextual theory in the same sense as the theories of Firth and Malinowski. The advantage is that the theory throws a bridge over the gulf that has existed between philosophical, sociological and

anthropological approaches to semantics. (Lyons 1977 II, 735).

3.3. Discourse analysis

A natural continuation of philosophical investigations into utterances ensues in the form of discourse analysis done by linguists, in which language is also viewed in terms of communication. Discourse analysis is concerned with the functional use of language. For discourse analysts the unit of analysis is also a speech act. Hymes (1972) insists that it "represents a level distinct from the sentence and not identifiable with any single portion of other levels of grammar, not segments of any particular size defined in terms of other levels of grammar." There are, however, certain problems in discourse analysis. Discourse analysts, for example, disagree on the number of functions, and the size of the basic unit. A difference between a philosopher's unit and a linguist's unit can be seen in the degree of idealization between raw data and the analyst's concepts. Philosophers concern themselves with decontextualized data whereas discourse analysts with raw data, which, however, are regularized and standardized to some extent, too.

In analyzing the features that link a series of sentences into discourse, two types of factors can be distinguished: cohesion and coherence. Cohesion is the fashion in which sentences and their parts are linked together to ensure the development of the propositional content. These links stretching over sentences are syntactic and semantic. A discourse is cohesive to the extent that it allows for effective development of the propositional content. The cohesion means that the linguistic forms are contextually appropriate. The coherence of the discourse, again, is recognized in the relationship which combines illocutionary acts through propositions. Propositions are not always overtly linked together. (Widdowson 1978, 27-29).

In the following examples the utterances are coherent but only the first dialogue is a cohesive text with the second sentence linked to the first by an ellipsis:

A: Can you fly to Oulu tomorrow?

B: Yes, I can.

A: Can you fly to Oulu tomorrow?

B: The Finnair pilots are on strike.

By sentences we express the content elements: a combination of propositions and illocutions. Sentences are linked together by cohesion as to form a whole which is called a text. A given series of sentences can make different combinations in which the proposition of the former sentence is connected with the proposition of the next sentence. The

illocutionary meanings have to be considered to obtain a good combination. The ways in which propositions are expressed and the ways in which they follow each other influence the illocutionary meaning. The decisive factor for a consistent whole of communication is the best possible combination of propositions and illocutions. A criterion can be found in coherence. "In the case of cohesion, we can infer the illocutionary acts from the propositional connections which are overtly indicated; in the case of coherence we infer the covert propositional connections from an interpretation of the illocutionary acts." (Widdowson 1978, 29).

Let us take an example of Widdowson's to illustrate how a piece of conversation can be understood by means of coherence although cohesion is absent:

A: That's the telephone.

B: I'm in the bath.

A: O.K. (Widdowson 1978, 29)

This dialogue can be interpreted coherent in the situation in which A's first remark is understood as a request to answer the telephone. Linguistically, the sentence is declarative, and not interrogative at all. In isolation, the remark cannot communicate the same thing, and in other situations, it might be used for other purposes. B's remark is a kind of excuse for not being able to answer the phone. A's second remark expresses an acceptance of the excuse, and his intention to answer the phone himself. Widdowson supplies the missing links to produce a cohesive dialogue:

A: That's the telephone. (Can you answer it, please?)

B: (No, I can't answer it because) I'm in the bath.

A: O.K. (I'll answer it.) (Widdowson 1978, 29)

What seems important to notice is that coherence is not only based on the linguistic code, although talking of illocutions and understanding the connexions between them may suggest it. As a matter of fact, illocutions are supposed to be included in the linguistic code, at least partly. Widdowson, it is true, states that the linguistic code only provides the clues for inference, and that we must inevitably rely on common knowledge. (Widdowson 1978, 31-32). What this leaves us with is that the linguistic analysis is so far too incomplete to provide us with an adequate description of the rules to be mastered in order to convey meaning by verbal communication.

The discourse rules necessarily raise the question of their origin. When the child acquires language, he also learns how the language generally operates. In the acquisition process of the formal system he also develops for himself an awareness of how linguistic elements are used in normal social interaction. He acquires a basic set of ground rules on

which communication as a common activity depends. The meanings of the sentences are not manifestly expressed but must be inferred. The child learns certain statistical probabilities of the frequencies of different linguistic elements, he learns certain common models of discourse development.

In addition to the above conventions, there must be certain nonconventional features which participants in conversation share in terms of discourse. They are certain conversational implicatures connected with general features of discourse. Our talk exchanges do not normally consist of a succession of disconnected remarks but they are characteristically rational and cooperative efforts for a common purpose. Grice (1975) labels such efforts as the cooperative principle. On the basis of this assumption he then distinguishes broad categories which he calls conversational maxims. The literal meaning of an utterance does not logically ensue a pragmatic implicature. It must be inferred from the cues of the utterance. The pragmatic meaning is based upon the utilization of the context, which makes of interpretation an inference process. It is done by means of inference rules. Such rules are to a large extent independent of language and universal features.

In conversation the roles of speaker and hearer change in turns, and besides, very little overlapping occurs. Conversation is characterized by a successive change of turns as speaker and hearer without pauses between them. Another characteristic is that the speaker roles are repeated. The problem lies in the fact that only one participant, not more, speaks at a time on the occasion. (Coulthard 1977, 54).

3.4. Linguistic context and situation

The words 'context' and 'situation' have occurred a couple of times in the preceding discussion without further definition. In this section, we attempt to define them more accurately. To recapitulate what we have analyzed, we can state that utterances are distinguished from sentences as follows: utterances are structures or units created in speech situations and dependent on language, whereas sentences are linguistic expressions used in speech situations. Sentences are linguistic units, consisting of linguistic expressions but utterances are sentences based upon the situation, i.e. pragmatic units. The context of given speech situations also consists of extralinguistic factors (the speaker's psychic state, his factual knowledge, his abilities, and other circumstances).

Situation and context are controversial concepts because they are used in various senses: either the situation is distinguished from the concept of verbal context or it is related to the general concept of context. The linguistic school of London (Firth, Halliday, John Lyons), who follow Malinowski, interpret context as a global concept including the situation. According to another interpretation the concepts are so far

from each other that no general concept covers them; context is always defined as a linguistic frame and situation as an extralinguistic frame. (Germain 1973, 23).

The classification from the psychological perspective in psycholinguistics (Slama-Cazacu 1961) and from the social perspective by Lyons (1963) leave the concepts too general. To facilitate a scientific description, Germain (1973) classifies less general types of situations. In this study we make use of Germain's analysis, and the context is confined to the linguistic environment, and the situation to the extralinguistic circumstances.

By situation is meant extralinguistic factors which are known to the speaker and the hearer at the moment of talking. Though the situation is theoretically defined as a function of both the speaker and the hearer, we must in practice content ourselves with the hearer's perspective. For the speaker's knowledge of facts, we should have to rely on introspection. (Germain 1973, 26, 29).

The factors known to the speaker and the hearer at the moment of conversation originate (i) from external physical reality (spatio-temporal), (ii) from nonphysical factors which are in the minds of the participants, (iii) from kinesic factors, i.e. body movements, or movements of a part of the body, facial expressions. Nonphysical facts are either psychological — the participants refer to things that they know of each other — or knowledge of geographical, historical, social, cultural, anthropological facts. At the moment of verbal communication they are then on the speaker's mind. (Germain 1973, 36).

The relationship between the context and situation resides in the fact that the situation can sometimes be of purely linguistic origin. The set of factors which are derived from the linguistic context and which the speaker and the hearer know at the speaking moment are called the contextual situation (*situation contextuelle*) (Germain 1973, 39—40). The term has the advantage that it defines the actual conditions of language use more accurately and that it can be employed abreast with the other situation types. This appears from the following scheme, in which the contextual situation is seen as different from the three other situation types.

— **CONTEXT:** linguistic

— **SITUATION:** extralinguistic,

a) of linguistic origin:

b) of nonlinguistic origin:

— contextual situation

— nonphysical situation

— physical situation

— kinesic situation

(Germain 1973, 44)

The different situation types, in addition to the context, can come up in oral use of language. In written language, we can employ only contextual and nonphysical situations. In the particular case in which

the text is on tape the hearer can, in addition to the linguistic context, resort to the contextual and nonphysical situations, just as in the written text. The kinesic and physical situations of the text cannot, however, be made use of.

By the analysis of the context and the situation, we have been able to isolate the extralinguistic factors of the communication situation which are part of the interaction but which do not reflect themselves in the linguistic code. On the other hand, the meaning conveyed by the language symbols, and the illocutions performed by the means of the constitutive rules, are revealed in the linguistic context or in the contextual situation.

To recapitulate, in this chapter we have distinguished important functions in the language code in terms of communication. We have observed that the meanings of the language are conventional to a large extent on the surface level. Despite the arbitrariness of the relationship between the linguistic sign and the referent there are obviously a lot of shared meanings among different languages, which enables mutual understanding of categories across the languages to such an extent that cognitive learning of foreign languages is possible. On the other hand, a fair share of the deep structure and functioning at that level is evidently common to different languages and their speakers, which is seen e.g. in the basic functions of illocutions and the development of discourse. A great portion of the meaning lies in the linguistic code. The meaning is inferred on the basis of the context and contextual situation. Although operation on the surface level is mainly done by means of the constitutive rules, it is possible to learn and understand another conventional system of symbols through the common ground of what it means to use language for communication.

4. Communicative Competence

4.1. Universal pragmatics

The speaker's psychic state, knowledge of facts, abilities as well as other extralinguistic factors related to the speech situation belong to the study object of empirical pragmatics. In speech situations are also repeated the general structures which are each time reproduced in employing certain kinds of linguistic expressions. These general structures which enable the speech situation are the study object of universal pragmatics, and belong to the theory of communicative competence. The task of this theory is to construct a set of rules by means of which are created situations which enable speech. (Habermas 1971).

The speech acts create the contextual situation, that is, the structures

of the speech situation. Simultaneously, these are revealed in the speech as linguistic expressions, which are called pragmatic universals (Habermas 1971). A speech act creates the conditions for the sentence to be used in an utterance, but at the same time it has the form of a sentence. The theory of communicative competence attempts to discover what the speaker or hearer does in transforming sentences into utterances by means of pragmatic universals.

In the utterance the speaker and the hearer use sentences in order to understand each other in regard to something. The double structure of the speech act consists of the combination of the performative sentence and the underlying propositional sentence, whether the performative is explicitly or implicitly expressed. (Habermas 1971).

To distinguish the communicative competence from the linguistic one, Habermas (1971) presents a three-stage abstraction process, starting from the concrete utterance. The concrete utterance is given in a certain situation and its meaning is also determined by the frame factors related to the situation: personality structure, and the role environment of the speaker/hearer. In the first stage of abstraction, the varying factors of the speech situation are ignored and only the general structures of the speech situations are preserved. As a result, we shall obtain an utterance which is the pragmatic unit of speech. When the communicative realization of the utterance is ignored in the second stage with the preservation of the linguistic expressions only, we shall have the sentence as the linguistic unit. In the third stage, maintaining the propositional sentence means omission of the linguistic elements which determine the pragmatic use of the sentence as the speech act. The basic unit left is necessary for the subject-matter to be interpreted, and in declarative sentences it is called an assertion.

Sentences are the basic units of linguistic study. As the theory of linguistic competence, the task of linguistics is — as also Chomsky (1965) sees the matter — to reconstruct a system of rules according to which a linguistically competent speaker forms and reforms sentences. Chomsky himself does not, however, abstract in the same way as Habermas. As a consequence, a direct, empirically verifiable (observable) relationship does not prevail between the ideal speaker and his concrete utterances. The discrepancy has been discussed by Itkonen (1974).

The basic units of universal pragmatics are utterances. As the theory of communicative competence, the task of universal pragmatics is to reconstruct a system of rules to describe how a communicatively competent speaker produces utterances from sentences. The concrete utterances are the study object of empirical pragmatics. Psycholinguists investigate the systematic variation caused by the dependency of the general structures of the speech situations upon the factors of personality structures. The theory of the linguistic code (sociolinguistics) studies the systematic variation which is caused by the dependency of the

speech situations upon the role structures. (Habermas 1971).

The following scheme displays the stages of abstraction:

Perspective	Theory
concrete utterance	empirical pragmatics (psycholinguistics, sociolinguistics)
utterance	universal pragmatics
sentence	linguistics
(assertion)	(predication logic)

Discussion of the logic is not needed here, but it completes the scheme.

The following catalogue enumerates the word classifications and their grammatical forms which refer to the general structures of the speech situations:

1. personal pronouns
2. words and phrases which are used to initiate speech and to address someone (grammatically: vocative, honorative)
3. deictic expressions (time, space); demonstratives, articles, numerals (grammatically: tenses, moods in certain cases)
4. performative verbs (grammatically: interrogatives, imperatives, reported speech)
5. nonperformatively used, intentional verbs, some modal verbs. (Habermas 1971).

These linguistic expressions are made pragmatic universals for the reason that they are related to the general structures of the speech situation: classes 1 and 2 to the speaker/hearer and the possible participant in the conversation; class 3 to the time, space and observational area of the utterance; class 4 to the utterance as such, to the speaker's relation with his utterance, and to the relationship between the speaker and hearer; class 5 to the speaker's intentions, attitude and expressions. (Habermas 1971).

Sentences cannot be used for utterances until the conditions which enable communication are created by means of pragmatic universals. The double structure of conversation requires understanding at two levels: at the intersubject level at which the participants can start a dialogue and thus act as capable of speaking and functioning, and on the other hand, at the subject-matter level, at which it is possible to describe facts. The pragmatically most important part of the speech act is the performative sentence. So far, there does not exist an accomplished system of the speech acts. As far as the speech acts are pragmatic universals, it must be possible to classify them universally, even though they are lexically complex in different languages. (Habermas 1971).

In addition to, or instead of the speech acts, there can be acts or emotional expressions (facial expressions) in communication. In

interaction all three phenomena occur together. The utterances are recognized as communicative as they are embedded in the extralinguistic expressions. Thus we can distinguish two forms in communication: the communicative action (interaction), and the linguistic discourse. The participants' actions and expressions occur abreast with the discourse but they do not constitute their material. In the discourse the participants try to establish the uncertain mutual understanding that has arisen in the communicative action. (Habermas 1971).

Habermas has attempted to characterize the ideal speech situation by means of the structural features of the situation that enable speech. The structural framework attempts to show what the communicatively competent speaker must do if he wants to participate in the discourse with the available speech acts. Habermas does not, however, believe that the conditions of the ideal speech situation correspond to those of the real speech situation. He outlines the structural frames of communication that can simultaneously illustrate the form of life if they are taken as the starting point of social theory.

4.2. The perspective of theoretical linguistics

Habermas' three-stage abstraction with the concrete utterance as the starting point provides us with the framework for the analysis of the concept of communicative competence. Competence refers to the ability to do something, but as the concept created by Chomsky (1965) it has been used in linguistics, from which it has spread out and gained a wider use. For Chomsky, the study object of the linguists was the linguistic competence of the ideal speaker/hearer which all native speakers of the language share. It means mastery of a very abstract and complex set of rules, by means of which the sentences of the language can be produced into concrete form.

The most important feature of the linguistic competence is creativity. It means the speaker's ability to produce new sentences which other speakers immediately understand though they have not heard them as such before. Since the speaker is a finite organism, the linguistic knowledge he has acquired is presumably limited. On the other hand, the language user is able to produce and interpret sentences infinitely. Therefore, it is possible to regard as an appropriate model for the speaker's linguistic competence a set of rules which itself is finite but which is capable of generating an infinite number of sentences for the speaker potentially to be able to produce and understand. According to Chomsky (1965) the linguistic description should tend to analyze the linguistic competence by defining the characteristics which are observed among the infinite number of well-formed sentences of a language.

The study object which Chomsky has presented for theoretical

linguistics is an ideal, completely abstract competence. And the same applies to the communicative competence created by Habermas. In this sense they are both equally abstract, but in the different stages of abstraction. Habermas does not assume that the model is necessarily realistic but he does not deny it being possible. The competence outlined by Chomsky constitutes the basis of the corresponding performance, but indirectly. Chomsky's notion is inconsistent at this point, he does not consider empirical investigations necessary for his hypothesis to be tested. Nevertheless, the native speaker can resolve the grammaticality of the potential sentences on the basis of his intuition because the competence is shared by the native speakers. This inconsistency has been shown by Itkonen (1974). Unlike his predecessors, Chomsky does not regard empirical sources — i.e. collections of speeches and writings — as an adequate source for the development of a linguistic theory, because they cannot possibly exemplify all possible structures. Therefore, Chomsky relied on intuition. (Coulthard 1977, 2—3).

In Habermas' stagewise abstraction, the linguistic description is put in its right place. Chomsky's own thinking does not place the model in that way because he is not able to organize the indirect relationship between the abstract set of rules and the observable linguistic acts (performance). Despite this inconsistency, the concepts of competence created by Habermas and Chomsky show equal abstractness on their own levels.

4.3. The perspectives of sociolinguistics and psycholinguistics

The linguistic competence cannot however be considered the idealized model for the linguistic performance (Chomsky 1963, 9), though the recursive feature of the generative grammar may suggest it (Derwing 1973). By the recursiveness of grammar is meant the characteristic that syntactical combination and submerging rules can be applied many times successively (Karlsson 1977, 33). Chomsky himself has emphasized that in principle recursiveness entails the possibility to explain the creative aspect of language use (Chomsky 1965, 205). Despite this central feature, the model lacks important features of linguistic performance. One of them is selectivity. A native speaker does not only produce new expressions but produces them on an appropriate occasion. Moreover, the context evidently greatly influences how the hearer interprets and understands the utterance.

Due to its narrow scope, wishes have been expressed to widen the competence suggested by Chomsky. Campbell and Wales (1970) argue that, in addition to grammaticality, competence must include the ability to use the utterances appropriately in terms of the context. According to Hymes (1972), Chomsky's concept of performance also lacks sociocultural meaning, and thus the social interaction is omitted. To

widen the original concept there has arisen the concept of communicative competence. Campbell and Wales (1970) view the concept from the perspective of psycholinguistics, and Hymes (1972) from the perspective of sociolinguistics. Both fields of study focus on empirical pragmatics: the concrete performance in a social situation.

Hymes (1972) wants to widen the linguistic description so that the speaking rules, manners, topics, forms of the message in various contexts are taken into account. He calls the field of study the ethnography of speaking. In the description of the speaking manner, four intercorrelating dimensions must be considered: (i) the available linguistic modes (styles), (ii) extrasentential structuring — such linguistic events as court trials, religious rites, debates, songs, (iii) interpretation rules by means of which a given linguistic construction gets its communicative meaning, and (iv) the norms of different interaction types. A description along these lines seems an impossible task otherwise than by confining to the subareas. Such a description can hardly provide much concrete use for determining foreign language skills.

Psycholinguistics studies language from the viewpoint of the individual in the reality of communication by a dynamic-contextual method and explains its use as a form of human behaviour (Slama—Cazacu 1976). The main objects of study are the linguistic performance and the acquisition of language. Sociolinguistics, again, is interested in the social factors that influence language use, with the focus on the society as a whole or on the individual as part of the society (Cripser & Widdowson 1975). Communicative competence thus implies a knowledge of the language in the same way as the linguistic competence. From this perspective, communicative competence means mastery of the linguistic code, and not pragmatic universals in the sense of Habermas' concept.

The communicative competence defined by sociolinguistics and psycholinguistics appears to be in conflict with the concepts of competence created by Habermas and Chomsky. With the help of their abstracted models they attempt to describe and understand the communicative and linguistic competence in order to find a system of rules, without interfering in the content, the language code. However, as a concept of empirical pragmatics, communicative competence structures the code itself, which shows that the term 'competence' is used in a different sense. An approximate equivalent might be an expression of skills or performance. If competence is employed as an abstracted term of empirical pragmatics, it should describe the particular system of rules which account for the systematic variation of the general structures between personality structures and speech situations (psycholinguistics) or between the role structures and speech situations (sociolinguistics). This use of competence would show an equal level of abstractness in the stagewise abstraction process from the concrete speech situation

upwards. The concept of competence would describe the original sense of ability, which again, might lead to a theory of performance.

Melenk (1977) attempts to restrict the concept of communicative competence by concretizing it and relating it to pragmatics. Following Peirce and Morris, he analyzes interaction and distinguishes the function of the utterance (pragmatic level), its contents (semantic level) and the grammatical-lexical structures and elements (syntactic level), by means of which the subject-matter is conveyed. These levels can be compared to three parts of competence. Communicative competence is confined to pragmatics, to the ability to act and react linguistically in the communication situation. However, it includes neither the ability to encode or decode the contents (competence of subject-matter, *Sachkompetenz*), nor the ability to form and understand correct sentences (language competence). Melenk resolves the problem of competence by dividing it into subparts which are partly overlapping and interwoven. He does not, however, characterize the concept at the same level of abstractness as Habermas and Chomsky in their analysis. This is only natural because Melenk's discussion is directed toward foreign language teaching.

As previously stated, Chomsky's concept of competence cannot provide a model for language performance. Neither can his performance model, though in it he parallels language use with performance. Chomsky ignores everything sociologically relevant, focusing only on psychological factors such as memory, perception, etc. (Hymes 1972). If we assume that the ultimate integrated performance model contains a single central, linguistic component common to both language production (encoding) and perception (decoding), the model must also possess the properties of bidirectionality. This means that it must be able to translate semantic representations into phonetic ones in production, as well as to interpret phonetic representations in terms of semantic ones in decoding. Transformational-generative grammar in all its different versions lacks this property. (Derwing 1973, 268—269). Of the three important components — recursiveness (or creativity), selectivity, bidirectionality — only the first component is included in the competence described by Chomsky. Consequently, it cannot offer a model for performance, which would be an important condition for a testable theory of language acquisition. Such a theory should account for how a preverbal child acquires the performance pattern.

5. Communicative Proficiency in a Foreign Language

The preceding theoretical background serves to illustrate and analyze language in terms of the native speaker when he uses it as a means of communication. The concept of communicative competence has been created to describe the abstract systems of rules for communicativeness and grammaticality which the native speaker unconsciously obeys. Pragmatics, which studies the code as concrete utterances, employs the term communicative competence for mastery of changes of registers. It was also shown that the different situations of language use display performance, in addition to the abilities. Since the conditions are not, however, the same for the foreign language learner as for the native speaker, the theoretical framework presented does not directly provide a model for foreign language learning. As a consequence, an attempt will be made to isolate the activities which are not necessary for learning a foreign language.

In his community man acquires the language which enables communication. As a member of the community, he adopts its paralinguistic forms which communicate attitudes and feelings and which vary in different cultures. Neither the paralinguistic phenomena, nor kinesics determine the meaning of the language (von Raffler-Engel 1976, Poyatos 1976). It is not necessary for a foreign language student to learn these ethnic features. It is unrealistic to think that he should identify himself with the native speaker, even though he must be able to understand and produce foreign language speech. Moreover, even if he wanted to master the foreign language completely, he would hardly ever be fully immersed in the foreign community linguistically (Corder 1973, 47). As for the unintentional communicative signals, we have previously excluded them from the definition of communication for this research purpose.

Man has learnt to make connexions by means of his native language. Thus the foreign language student already has one linguistic means of communication. In the use of the vehicle, all languages embrace universals which ensure interaction at the linguistic level but which are not, however, necessary for foreign language learning: predication, reference, persons, time, space, deictic expressions, questioning, commanding, etc. Such phenomena in conversation as change of the roles of the speaker/hearer, anacolutha, repetitions, hesitations, ellipses, contaminations, etc. can also be included in the universal features.

The different languages also share the basic grammatical organization: nouns, verbs, adjectives, syntactic relations, etc. Though the classification may be differently expressed, the organization itself is similar. The grammatical organization in the brain is part of the limbic system, which has phylogenetically developed prior to the nonlimbic

cortex system. The child's development of grammaticality is almost accomplished by the age of five (Walz 1976). Wittgenstein's 'pure crystal' (1953, § 97) may also refer to this *a priori* order of the world.

The coherence and cohesion of the discourse are maintained by the global rational principle which enables cooperation in conversation with various anticipations. In order to understand the subject-matter completely, the participants must be able to infer by means of different cues. Such inferences are universal and global in nature. Cohesion and coherence are used to generate linguistic context, which as a whole constitutes a discourse. Thus the development of discourse is a universal feature.

Since the preceding universals are acquired in the acquisition process of the first language, they can be considered abilities or competence. The concept 'competence' has been used to demonstrate the scientific models and theories by means of which the native speakers' linguistic abilities are accounted for at an abstract level. In the learning process of a foreign language these abilities can no longer be acquired. As a consequence, it cannot possibly be justified to regard communicative competence as the goal of foreign language learning.

The symbols of the language are used according to certain rules. They are constructive rules, which are observed intuitively. They can be said to be part of man's linguistic sense, to constitute the linguistic competence. For a foreign language learner, the rules are not constitutive but regulative rules, because they are not mastered intuitively. As a consequence, it is not justified to speak of a foreign language learner's linguistic competence in any other sense than to refer to his mother tongue.

The foreign language student must learn new signs, symbols and their meanings. Since the symbols vary in abstractness, we can think that some symbols can be learnt if they are ostensively defined. Nevertheless, an apparently concrete object or thing does not always have the same scope of meaning in different languages. The indication of the range of meaning can often be difficult which therefore calls for a wider context to infer the meaning of the symbols. Consequently, the foreign language must be presented in context. In various linguistic contexts the scope of the meaning can be specified. We can also suppose that part of the various sociocultural, historical, and geographical factors are revealed in the linguistic code. Nevertheless, on the basis of the concept of the nonphysical situation we may assume that there is a certain amount of the background knowledge which does not explicitly come up in the linguistic code though it is presented in the context. For that reason, some factual knowledge is apparently needed to facilitate efficient use and understanding of the foreign language.

The foreign language is learnt to communicate with the speakers of the language. For the communication to succeed orally, it must be

performed fluently enough. Fluency is a complex and difficult concept to define. Its analysis is, however, important for the determination of the level of language proficiency. Perhaps the exact same criteria of fluency cannot be applied to the written discourse, but an attempt must be made to determine them, too. In addition to fluency, proficiency in a foreign language comprises mastery of various contexts. In the dimension of abstractness, we might suppose that it is more difficult to master a context, the subject-matter of which is abstract than the subject-matter of a concrete one. To a foreign language learner, the difficulty of the context can also appear in the strangeness of topics and concepts. The scope of the subject-matter could then be another dimension of mastery of the context.

Generally speaking, the preceding discussion may justify defining the student's communicative proficiency in a foreign language as oral and written fluency in various linguistic contexts. It would be language use. The term proficiency is preferred to competence because the fluent use of foreign language symbols in context and discourse is a proficiency that has been gained through learning. It is not an inborn faculty or an ability that has been acquired rather unconsciously. Moreover, the term 'competence' for educational use can be criticised on logical grounds. Pearson (1980) has analyzed the concept and come to the conclusion that competence suggests the mastery of basic knowledge in the field, a kind of minimum level. "Competency does not describe a level of performance somewhere between just knowing and knowing very well how to do something. Rather it indicates the level at which a person's performance becomes commendable" (Pearson 1980).

The term 'competence' also suggests its relatedness to the behaviouristic view. Roughly speaking, skill knowledge can be divided into two general classes: habitual skill knowledge and intelligent skill knowledge (Pearson 1980). The former refers to routine performance without much reflection and the latter to performance that requires insight, understanding and intelligence. The rationale of the behaviourist view rests upon the idea that intelligent skill knowledge can be reduced without loss of meaning to habitual skill knowledge. In terms of foreign language learning it is possible to confine the educational goal to the mastery of basic knowledge and speak of competence. But for a more general meaning the term is too restricted. Within the scope of a broader sense of the term, it is possible to specify various learning aims. An advanced knowledge of a foreign language can hardly be transformed into mere behaviouristic units. Although there are aspects in the use of the language that require routine behaviour, on the whole, such a performance calls for insight and understanding, too. Consequently, the term proficiency can best describe the knowledge that is needed for communication in a foreign language.

An interpretation of language proficiency for communication that is

intentionally quite similar to our definition is advocated by Widdowson. His term 'use' to be distinguished from 'usage' describes it, but he does not reject conceptually the concept of communicative competence (Widdowson 1978, 67, 144). In different views presented there occurs variation in emphasis but, on the whole, they are conceptually associated with communicative competence. It is unnecessary to report them here because they rely on the same sources and interpretations which have been presented in theoretical linguistics and sociolinguistics.

The following chapter will discuss the instructional application of communicative proficiency in a foreign language within the pedagogy of foreign languages. We shall outline an instructional model, by means of which an attempt is made to deduce the criteria for evaluating communicative proficiency in a foreign language.

6. An Instructional Analysis of Foreign Language Proficiency

6.1. Foreign language pedagogy¹⁾

Foreign language pedagogy is not a study based on a unified theory. It is still a controversial and ambiguous concept and field of study. Its relatedness to many sciences with their mutually different theories and concepts makes the field difficult to organize as a branch of study. However, foreign language teaching is such a complex question that all the relevant aspects should be taken into consideration under a single discipline²⁾, particularly, as it is most often concerned with organized education.

Traditionally, the problems of foreign language teaching have been dealt with on the basis of language analysis and psychology. Linguistics has been applied to the language part of 'language learning', and psychology to the learning part of it (Mackey 1973). As a consequence, the principles of the one or the other have been in prominence as a cyclic shift. Today linguistics is emphasized in the form of applied linguistics.

There is at present no single comprehensive theory of language. That is why language may be studied from different perspectives: in

¹⁾ The word for word translation is 'foreign language didactics'. The term 'didactics' is used to describe 'the science of teaching' without any authoritative connotation. 'Pedagogy' is preferred to here because didactics has a pejorative nuance in English. The concept slightly suffers from the translation equivalent.

²⁾ By discipline is meant a branch of learning at university, not a discipline which is autonomous in its contents like mathematics, physics, etc.

sociolinguistics as a social phenomenon, in psycholinguistics as an aspect of the behaviour of the individual or in theoretical linguistics as an object itself. Unlike sociolinguistics and psycholinguistics, which are intimately connected with sociology and psychology, respectively, the study of language structure has developed its own theoretical framework and terminology. Hence the notion of linguistics as an autonomous discipline. Applied linguistics makes use of all these theoretical approaches. (Corder 1973b).

There is some confusion about the terms of applied linguistics, and linguistics applied to the teaching of languages, because they are both called 'applied linguistics'. There are two types of combined disciplines: (1) those in which linguistics is applied to another discipline (translation, lexicography, ethnology, teaching of languages); (2) those in which another discipline (e.g. statistics, logic, mathematics, etc.) is applied to linguistics. Moreover, a distinction is not made between linguistics applied to language teaching and methods of language teaching. (Galisson et al. 1976, 35).

Along similar lines, Malmberg (1971) regards applied linguistics as a science¹ with primarily two perspectives: (1) linguistic theory as a basis for different social activities and necessities, of which language teaching is perhaps the most important, but not the only one; (2) the scientific analysis and testing of means, methods, procedures, and policy within the different fields of education, therapy, pathology, administration, social and cultural activity. Although these two perspectives overlap, they do not coincide.

Since applied linguistics is concerned with language teaching only to a limited extent and ignores the methods of teaching leaving them to pedagogy, it is justified to deal with language teaching problems in a more compact scope, and thus to develop foreign language pedagogy as an independent field of study, as a branch of the science of education. Pedagogy is not limited to instruction, but views education from a larger perspective, so that teaching of foreign languages appears in a broader frame.

As an independent field of study, foreign language pedagogy, of course, synthesizes results of suitable disciplines; in fact, the relevant findings in the appropriate fields are necessary for the development of foreign language teaching (Mackey 1973). Nevertheless, being a branch of education helps foreign language pedagogy to achieve such a stability that unnecessary fluctuation entailed by new steps of language analysis or learning theories is avoided. As Malmberg (1973, 25) has pointed out, every phase in the development of a language theory has corresponded

¹By science is meant any branch of the knowledge that has accumulated on a subject, and not only natural sciences but also humanities, social sciences, etc.

to a new phase in language learning theory.

The connexion of foreign language pedagogy with its contact fields might be described as in Figure 2. It is to be noted that school reality is included in the figuration. This aspect is omitted in the French definition of foreign language pedagogy (Galisson et al. 1976, 151). School reality may suggest that foreign language pedagogy is limited to school education. This need not be so, although it is obviously the main field of interest.

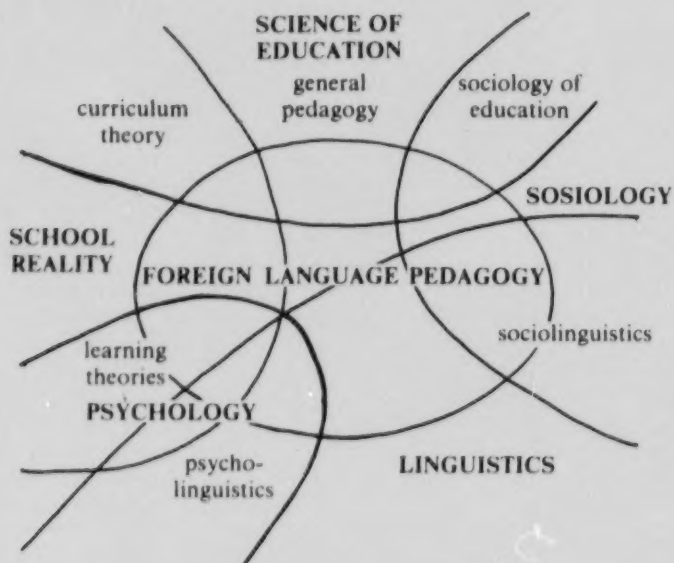


Figure 2. Foreign language pedagogy with its contact fields according to Leino (1979, 16).

Being a branch of educational science, foreign language pedagogy has to face the problems of theory development within educational research. The main issues of teaching are concerned with the conceptual definitions of and the relations between general pedagogy, subject-

related pedagogy, curriculum, curriculum theory, and the question of whether the disciplines are normative or not. In the following the problems will be dealt with only to the extent necessary for the development of the subject.

Pedagogy is used as a scientific term in education to cover all theories of teaching which conceive of and account for teaching in regard to aims, contents, premises and reasons. In a wider sense of pedagogy, methods are included as theory of teaching procedures, which makes pedagogy the core of an overall theory of school (Wörterbuch 1977, I, 193—194). This definition is given of general pedagogy, from which different types of pedagogy can be separated, subjects to be taught being the commonest criterion.

Roughly defined, curriculum has been used in a broad sense as a plan for the whole educational system, and in a narrow sense, including only contents. In the former sense, the term 'curriculum design' might be used. Such a curriculum design comprises a body of all major arrangements and measures to be taken within the frame of the set goals in order to provide pupils with conditions for favourable learning experiences (Koskenniemi & Hälinen 1970, 106). Such curriculum designs can be of different types: subject-centered (Lehrplan), individual-centered, or society-centered. The underlying theoretical base rests upon curriculum theory, which guides the selection of curriculum aims, content and method. Choosing the subject-centered design as the structure of curriculum enables creation of the different subject-related branches of pedagogy. Within the subject pedagogy, the content and method of teaching as well as their interrelation are defined in the subject curricula.

The controversial relation between pedagogy and curriculum is mirrored in the question of the normativeness of pedagogy. The issue is actually intertwined with two aspects: on the one hand, norms and values approved of in a society and their influence on pedagogy, and on the other hand, the norms and prescriptive rules given by pedagogy. These two facets, it seems, are not always recognized although they are embraced in the concept 'normative'.

In a scientific sense, normative pedagogy is to be understood as a system in which binding educational aims, contents of instruction as well as methods and educational forms are deduced from general pre-pedagogical norms (Sinn-Normen). Normative pedagogy presupposes that the higher norms already incorporate all pedagogic realizations. Such a conceptualization is not scientifically tight because pedagogy is conditioned by the kind of factors that cannot be derived from the higher norms. (Blankertz 1975, 19). Therefore, very different conceptions and interpretations are plausible within a given norm system. Because of this, the scientific grounds of pedagogy differ from those of the basic research in education. As Malinen (1974a, 32—33)

points out, the starting point of pedagogy is intentional action: the action-related practical results are an essential part of pedagogy. As a consequence, a pedagogic theory based on aims and intentional activities cannot be a static survey of what occurred but must be concerned with some intentional action.

The other facet of normative pedagogy is concerned with the question of whether pedagogy can provide directions and prescriptive rules. A partial answer to this can be concluded from the above cited remark by Malinen. The practical research results are to be made use of in school reality. Lahdes (1977, 34) also considers it necessary for pedagogy to facilitate teachers' work with guides and rules about how to put the official curriculum design into use. In the same category with normative pedagogy Blankertz ranks compendium literature (*Kompendien-Literatur*), with which he refers to apodictically formulated instructional guides and advice, definite procedures and opinions given as binding norms. He condemns such literature as not belonging to systematic pedagogy. (Blankertz 1975, 18). The main reason for his argument is that such guide books are unreasonable as isolated and extracted out of the context of theoretical foundations.

The above discussion entails the framework of how to define the mission of foreign language pedagogy. The question concentrates on instructional curriculum. A subject-related curriculum in foreign languages can be justified by them being the objective as well as the means of instruction (Leino 1978, 11). Therefore, the mission of foreign language pedagogy can be defined as to develop practicable curricula of foreign languages with the related linguistic objectives, instructional contents and methods, and to investigate their realizations in practice. (cf. pedagogy of mathematics, J. Leino 1977, 10). This definition differs from that of Lahdes (1977, 34) on general pedagogy, in which he only concentrates on methods. The development of curricula can be regarded as central because the content and form of instruction and their interrelation are intertwined in the curriculum, which, again, becomes prescriptive and thus normative for instruction.

As a rule, subject-related pedagogy follows the structure of its subject discipline (Koskenniemi & Hälinen 1970, 186). This does not apply to foreign language pedagogy offhand. The elements analyzed by linguistic science cannot be formulated as learning objectives. Instead, linguistic description may be used to describe the learning objectives. In addition, decisions upon selection, gradation, and presentation of learning objectives cannot be made just on the basis of linguistic description. Language learning is not merely understanding and remembering of knowledge but, to a large extent, an instrumental subject that demands a lot of training and automation at all levels.

6.2. A sketch of an instructional model for foreign languages

In the preceding chapters an attempt was made to isolate those factors which are mastered in the first language. Nevertheless, it does not follow that they do not accompany foreign language learning. The analysis was to show that those factors are no longer acquired in learning a foreign language but that they form the basis on which the new material must be anchored and by means of which a language is organized and learnt for communication. Although the rules of a foreign language are cognitively understood, a lot of practice is needed for the automation of the new signs.

Roughly speaking, we can view the meaning of language in terms of symbols and the intentionality conveyed by them. A foreign language student must learn the meaning and form of the language symbols and their interrelated order in the surface structure. This whole consists of phonological, orthographical, morphological, syntactic and semantic factors, all of which together can be called the object language. In communication with other people a foreign language student must also be able to express his own intentions and to understand others' intentions. The intentional meaning is revealed in illocutions, which again are combined into a relational network by means of coherence and cohesion. The totality of such a network constitutes a discourse: a letter, a report, an article, an account, a description, a complaint, a speech, etc. The whole made up of illocutionary meanings can be called the discourse language.

The construction that holds together the object language is the grammar of the foreign language. Since it does not stand for the whole grammar of a given language within instruction but an applied description of the rule system for the foreign language student, it may be called an instructional grammar or a pedagogic grammar. The discourse language probably cannot be structured into a rule system similar to that of the object language. To a foreign language student the system of the object language stands for regulative rules. The uniting principles of the discourse language are, however, to a large extent global, and thus universal. Therefore, they are obeyed rather unconsciously and can perhaps be compared to the constructive rules. In the communicative use of the foreign language these rules come into an instrumental position whereas the rules of the grammar must be followed knowingly, at least until they have become sufficiently automatic.

Using the discourse language as a communicative medium can be conceived of as processing discourse. Then it also embraces the object language. Accordingly, we can operationalize communicative proficiency in a foreign language into the skill of discourse processing. In terms of communication, this implies processing oral and written discourse as both a sender and receiver of messages. The discourse

processing is a term to describe the level which the student must have reached in order to be able to communicate in the foreign language. However, it does not describe proficiency or performance in absolute terms, but rather expresses that the foreign language is used for transmitting messages.

For instruction a third variable is needed: the metalanguage. The metalanguage is used to discuss the object and discourse languages. The instructional operationalization of communicative proficiency may be described as a discourse processing model which is visualized in Figure 3.

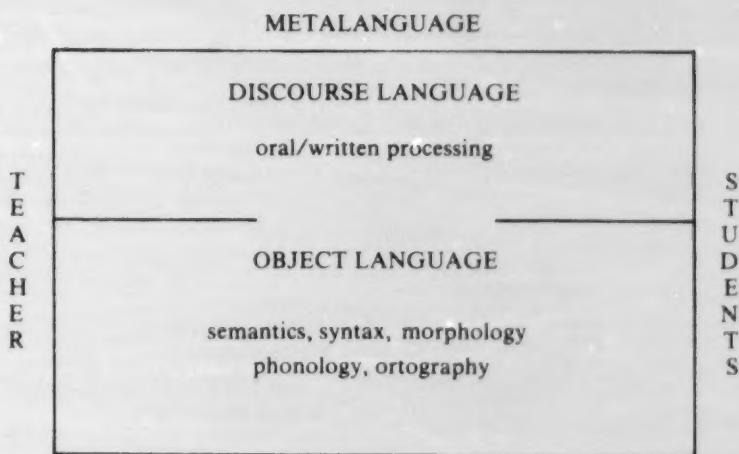


Figure 3. An instructional model for foreign language teaching.

Instructionally the model is an interaction event, which consists of the teacher and his instructional management and the students and their activities. Since the model entails a third component, the metalanguage, it is possible to confine to the teacher component all the teacher's educational and instructional aims and intentions other than those that belong to the metalanguage. The same distinction applies to those activities of students' that are not part of the metalanguage.

The metalanguage is essential for instruction because by means of it the teacher describes, explains and presents the object language and the discourse language to the students. It also describes how the students react to the discussion of and how they function in regard to the object and discourse languages. The first thing one comes to think of as

constituting the metalanguage is the choice between the mother tongue and the foreign language. This can be said to be one of the main components of the metalanguage. The other component consists of techniques. It is possible that the teacher wants to transmit the rules of the object or discourse languages, and meanings of words or phrases through demonstration, visual aids or graphically, without using any spoken language at all. This would, however, be using metalanguage in a broader sense of the term. The component of techniques forms a large and complex area, which is here left out of the discussion but which needs further investigation. On the other hand, the techniques must be distinguished from other instructional activities, which were included in the teacher component.

The metalanguage as spoken or written signals provides four possible communication solutions between the teacher and the students. The teacher uses the native language as the metalanguage and the students react either in their mother tongue or in the foreign language. The other two alternatives are that the teacher uses the foreign language and the students either their mother tongue or the foreign language. When the foreign language is used as the metalanguage by both the teacher and the students, the situation arises where the metalanguage has reached the level of discourse language. It is only a matter of perspective if the distinction is kept to emphasize that the topic of discussion is the object language and not some other general subject-matter. The following scheme displays the alternatives of using the metalanguage in instruction (FL = foreign language, MT = mother tongue):

Teacher	Students
FL	FL
FL	MT
MT	FL
MT	MT

In the use of the discourse language the same alternative cases occur between the teacher and students. When the participants do not speak the same language, discourse processing is incomplete and a temporary instructional activity.

The last thing about the model is to remind the reader of its function. It is built up to describe the essential components in the form of a model for instruction. It is theoretical to such an extent that a bridge will have to be built between it and the empirical phenomena, which calls for further research on the problems involved.

Instruction and learning must start from the object language. The model must be understood so that the student advances from the level of the object language to that of the discourse language, even with a limited amount of the language material. According to the model, advancement takes place like in making a rope. The strands are twisted together with

binding rules until a string is formed. These strings can be conceived of as illocutionary acts which gradually form a discourse cord. Advancement can be seen in the process of a cord transforming into a rope. The comparison shows that from the very beginning the aim is to use the language as a vehicle. During the process the vehicle supersedes the constituent elements of the object language just as in place of the material of the rope, attention is paid to its potential use.

Applying and modifying the definition of learning given by Heikkurinen (1971, 126), by learning the object language is meant formation of a new cognized concept through a thinking process within a person's consciousness, which results from the immediate interpretations of the object language and/or from a combination of meanings related to the previous or simultaneous descriptions of the object language.

For foreign language learning the previous meanings are often derived from the mother tongue, especially at the beginning of study. The scopes of a meaning are often the target of the description. Later on, at a more advanced level, meanings can be precisions or extensions of previous ones. When a person has learnt more foreign languages, the combination of these meanings into a new foreign language can be a factor that facilitates learning, particularly if the foreign languages have common features.

Since the foreign language learner must also be able to apply the concepts of the object language, he must know how to process discourse. In this sense discourse processing can be considered the phenomenon of transfer because the student is made to employ his cognitions instrumentally and in new situations. The distinction between learning and transfer is characterized by time interval: learning refers to a temporarily shorter period than transfer (Heikkurinen 1971, 129).

The mission of foreign language pedagogy is to develop appropriate curricula which define the contents, and the methods by which the contents are studied. It is worth noting that the method is used in a broader sense, and not in the sense of the teacher's techniques. The discourse processing model serves to give a curricular description. In it we have attempted to develop a practicable curriculum for the aim of communicative proficiency in a foreign language. Favourable learning experiences means learning which is then applied in the form of transfer. On this basis a student is assumed to be able to use his language beyond the instructional situations.

As a matter of fact, the model is a curricular operationalization of the set educational goal, which makes it instructionally normative. Accordingly, the realization of the curriculum can be evaluated e.g. by assessing learning experiences. The model also serves to develop models for evaluating foreign language proficiency, as well as to determine various syllabuses if needed.

The model is narrow in scope because foreign language instruction is not placed in a wider educational context. This need not be a problem, rather it can be an asset. The model does not prevent us from utilizing e.g. psychologic research or different learning theories as well as linguistic studies. The model can also be integrated into general pedagogy and educational aims because they can offer the basis for the choice of the subject-matter, the transmitting of which is desired in the foreign language instruction and by which the consciousness of the students is influenced. Further research can thus make use of this model and integrate it into a broader pedagogical model or theory.

The teacher and students components are not analyzed in this study. They however embrace important factors to be taken into account: personality, abilities, interests, motivation, etc. Since such factors have not been described in the model, it offers different solutions for further research. Despite its narrowness, the development of the model can be justified by the importance of clarifying the linguistic elements necessary for foreign language instruction. It is also an advantage that it is not static and that it is not bound by theories of theoretical linguistics. A more serious weakness is that the model does not give solutions e.g. to gradation and sequencing of the language material. Nevertheless, such basic assumptions, dependencies, and testable hypotheses are required of a theory. The model does not take into account interaction between students because the instructional arrangements are thought of as performance of the teacher even when students work in pairs or groups.

6.2.1. An instructional syllabus of foreign languages

By syllabus is here meant the contents of the foreign language teaching programme. The syllabus includes the grounds for the choice, structure, and presentation of the content. Various taxonomies have been used to classify the content material. The base for the categories has been laid either on a psychological, hierarchical system (the Bloom system by Valette 1971) or on the grammatical classification. The foundation of the hierarchy in the psychological classification lies on the assumed order of learning. Originally the taxonomies were designed to facilitate evaluation of learning (Leino 1978, 20). The linguistic research has provided the traditional grammatical model for categorization.

The emphasis on the functions of the language has contributed to the creation of situational syllabuses. They attempt to meet the demands of the learner by answering the questions of when and where the student needs the language that he has learnt. Categories and emphases in the classification have varied to some extent, but they share the fact that they apply best to the so-called general education (Trim 1971), to everyday and restricted language use (van Ek 1976) or to foreign

languages for special purposes (Howatt 1974, Munby 1978). The focus of these syllabuses is on the student and primarily on his needs.

Wilkins (1976) has developed a notional syllabus which is appropriate for communicating thoughts. In it the linguistic content is planned according to the semantic demands of the learner. The concept of communicative competence has influenced the construction of syllabuses. The different communicative functions have served as the basis for categories (Piepho 1974b) and the sociolinguistic theory by Hymes (1972) has exacted the classification. The following discussion will deal with the presentations of Wilkins and Piepho because they can also be applied to school teaching.

As a result of the application of the criteria, Wilkins (1976) distinguishes three kinds of syllabuses: (i) grammatical, (ii) situational, and (iii) notional syllabuses. The progress criteria of the grammatical syllabus are simplicity, regularity, frequency and contrastive difficulty. (Op.cit. p. 6). In the situational syllabuses an attempt is made to predict the situations in which the student is likely to need the language and then to teach the language necessary for those situations. The analysis of the situations tends to take into account all the relevant factors, such as the channel, the context, the roles, and the relationships between the participants (Op. cit. p. 16). These syllabuses progress linearly but do not pay enough attention to the communicative capacity. With it as the starting point Wilkins develops the notional syllabus for the student's needs to communicate his thoughts. (Op. cit. p. 19).

Wilkins (1976) distinguishes three types of meaning that can be conveyed in the uttering of a sentence. The cognitive meaning (Halliday's ideational meaning, or propositional meaning) is expressed through grammatical systems in different languages. The speaker's attitude to what he is saying is expressed by the modal meaning. The third type of meaning is a matter of the function of the utterance as a whole in the larger context in which it occurs. This is labelled the functional meaning. The semantico-grammatical categories express concepts or conceptual meaning: time, space, quantity, matter, deixis and case. The communicative functions are divided into six types: judgement and evaluation, suasion, argument, rational enquiry and exposition, personal emotions, emotional relations. (Wilkins 1976, 44—54).

The functional and conceptual categories together form the notional classification. The inventory of notional categories are intended as a tool in the construction of syllabuses. The inventory is rational but not hierarchical, and its categories are partly overlapping. (Wilkins 1976, 24, 42). The notional syllabus lends itself well to a cyclic approach. From the simplest and least differentiated manner of expressing himself the student progresses to the semantic repertoire of the most advanced level. (Wilkins 1976, 59).

Piepho's analysis of the aim of communicative competence in English (1974b, 9, 82, 84) has its foundation in Austin's (1962) and Searle's (1969) theory of the speech acts. Piepho replaces the principle of the linguistic progress by the immediate functional needs and the student's speech actions. The classification is obviously based on the speech acts, in relation to which Piepho organizes the linguistic content (Piepho 1974b, 10).

Wilkins's notional syllabus implies a certain connexion with the speech acts theory. The semantico-grammatical category corresponds approximately to the propositional acts because the deictic class expresses reference, and the other belong to predication. The communicative functions can be compared to the illocutions. (Mindt 1978).

Wilkins's semantico-grammatical categories are parallel to the communicative categories, but Piepho (1974b) classifies the concepts as formalized concepts of the semantico-grammatical category, and distinguishes them from the psychologically and sociologically determined intentional actions and functions. As a result, the parallel to the speech acts is no longer valid, and the views differ as to the importance of the meaning of the linguistic elements to the illocutions (Mindt 1978). Piepho gives preference to the illocutions (1974a, 10), and only secondary are the linguistic aims (1974b, 78). As a consequence of this view, Piepho does not succeed in integrating the grammatical material with the situational areas (Mindt 1978), whereas Wilkins (1976, 56—57) reminds us that there is no one-to-one relation between grammatical forms and either grammatical meanings (conceptual meaning) or language functions. Therefore the familiar grammatical categories must be integrated into the semantico-grammatical category.

Wilkins's syllabus has been criticized because the classification has been made rather intuitively (Leino 1978, 22), and also, because in the system attention is paid to separate units — utterances usually — without bringing forth the concept of the discourse structure (Coulthard 1977). There are other weaknesses, which Wilkins readily admits himself, but nonetheless, the syllabus can be considered an interesting attempt to structure the content material in a way other than by relying on the grammatical syllabus only.

It is not very likely that the content aims of the language can be classified by means of the illocutions because opinions differ drastically about their number. Austin (1962), for example, estimated their number to be as many as 10 000, which he classified into five categories. Searle (1969) does not estimate the number but needs many more than five classes. Some discourse analysts think they need 22 categories (Coulthard 1977, 25). Nevertheless, the classification of the illocutions does not provide a hierarchical system. They do not offer clearcut categories because the classes of the performative verbs and indirect

performative expressions would have the same elements in the different classes, and thus cause overlappings. The large and various aggregate of illocutions does not serve as a useful and economical organizer for the language. Neither does it guarantee the occurrence of all the necessary grammatical forms. This is due to the fact that the same regularity which exists between the form and the linguistic system does not prevail between the linguistic form and the function of the language use.

The grammatical classification, however, offers us a more concise framework for categorizing the language recursively. The syntax provides a cohesive network for a more detailed analysis of the content material. It is more logical to start from the narrower concept of predication-reference and then progress to various acts and attitudes than follow the inverse order. The instructional arrangements in schools cannot possibly cover such specific situations that they would anticipate all the coming circumstances that the students will later face. Therefore, it is necessary to learn a construction which enables generalizations over situations (transfer). If various situations are elements, the knowledge of which enables the building of hierarchical and more complex networks, the very situations should then become more complex too. Obviously it is not so. Instead, from the one-word sentences the student may progress to more complex sentences and utterances, and eventually learn to master the complex language. Therefore, this intralinguistic classification can guarantee the framework necessary for the transferring use of the language.

Although the starting point is in the basic structures of grammar by means of which the student can reach generalizations in the mastery of the language material, it does not follow that the whole set of rules must be learnt first. According to the instructional model for discourse processing, even the elementary rules are sufficient for the student to advance to the functional use of the language. Thus the problem ultimately focuses on the choice of such material that offers an integrate progression of the grammar and functions of the language.

Hymes's (1972) sociolinguistic theory, which tends to consider grammaticality, psycholinguistic realization, sociocultural appropriateness and the actual performance, is not sufficiently organized in detail to provide the criteria for the choice of the linguistic content (Leino 1978, 62). Utilizing the sociological theory, Halliday (1973) approaches the language and the speaker's use of language sociosemantically. He however warns against an excessive generalization, because on the intellectually high-level, abstract discourse — and in general, when the language itself creates its own context — it is not possible to describe the language in sociological terms.

This view is strengthened by Williams's (1970) classification of the modes of speech into broad categories where there are intersections of particular patterns of language forms with particular functions in speech

situations. They are as follows:

1. *Impulsive*. Utterances that have very minimal linguistic forms, often single vocalizations with no syntactic pattern requirements. Their function is usually a speaker's vocal reaction to a situation such as crying "ouch," or laughing, crying, swearing, and the like.
2. *Contactive*. Again, utterances are minimal, often word forms where syntax is typically minimal. Here utterances reflect a sender's attempt to initiate or maintain attention with another person, and topic is irrelevant. Examples are "Hello," "How are you?" "Do you hear me?"
3. *Conversative*. Minimum word forms are typically used; however, syntactic fragments also appear. Utterances serve to reinforce verbal interaction with another person. Topic is not of ultimate importance but is relevant to maintain the exchange. Examples are cocktail party chatter, elaborated greetings and farewells.
4. *Descriptive or directive*. Minimal word forms and syntactic fragments are possible, but syntactic elaboration is usually required. Speech is either in the first or in the second person. The function is to *describe an experience* where the organization of the description follows the same sequence as the experiences, or *direction-giving* where verbal organization is closely related to the sequence of behavior which is prescribed.
5. *Elaborative*. This mode requires syntactic patterning, often elaborated, and overall message organization usually includes some major structure. Communication of the topic to the listener is the prime function, where explicit topical elaboration reaches to levels which could only be obtained through verbal symbolism. Examples are argument, the interpretation of some event, relatively complex explication, and the like. (Williams 1970).

The classification is general; e.g. doing business at the post-office or bank, enquiries, etc. are not expressed, but they evidently belong to category 4. The advantage of the classification is that it shows on what basis foreign language materials may be determined. Categories 1—4 are largely based on situations, but even they must embrace the grammatical structures as systems, not as fragmentary units. Moreover, the material must include prose text. In the last category the linguistic material is more dependent on the intralinguistic mechanism than in the other categories. Therefore, it lends itself well to authentic texts.

At the university of Nancy, France, experiments with the authentic material have been made from the elementary level onwards. However, they have had to make such alterations as re-recording to improve the acoustics or to eliminate the idiolectic features of hesitations, repetitions, errors. (Riley 1976). In the structuring they have had to separate the teaching of the functions from that of the morphology and syntax. (Riley 1976). Jakobovits (1970) also distinguishes the communicative aims from the teaching of the grammar. The model developed in this study for discourse processing also isolates the aims in the object language from those in the discourse language. Despite the drawbacks and difficulties, authentic texts, through some normalization, could perhaps serve as appropriate material for language teaching, which would be functionally genuine.

Foreign language syllabuses for advanced levels, like for senior secondary schools, might be planned by means of contexts. They could possibly facilitate determining the scope and abstractness of the subject-matter. In theory, variation in dialectal and social variants belongs to the whole domain of the language, out of which a sample might be e.g. the so-called standard language. By it is meant the form of language by means of which people manage the everyday activities of life in society. The language of the mass media is primarily of the standard language, which is also used by the public administration in their communication with the citizens. Also, most laws and acts are written in the standard language. Generally known expressions of the language are used in the standard language. (Kielikello 10, 19).

The above definition is broad and leaves a lot to be interpreted in the choice of the contexts. Such a definition does not, however, give time-bound criteria for the language material, like the classifications made by means of glossaries or frequencies. The process of change is, in the semantic meaning, more rapid than in the features of phonology, morphology, or syntax. Besides, the frequency lists may be misleading if the other aspects of the vocabulary are not considered, such as range, availability, familiarity and coverage.

6.2.2. Analysis of foreign language skills

6.2.2.1. Foreign language skills in terms of communication

Widdowson(1978) makes a distinction between language usage and language use. The first of these refers to the knowledge of words and sentences as manifestations of the language system, and the second is the way the system is realized for communicative purpose. The distinction may be paralleled with the difference between the object language and the discourse language. At the former level the knowledge of the language is mastery of the language system and at the discourse level mastery of the language use.

Traditionally, the four skills of language proficiency are speaking, listening comprehension, reading and writing (Lado 1961, 25; Harris 1969, 9-11; Nykykiet 1971). These have been classified in various ways. Speaking and listening have been related to the aural medium and reading and writing to the visual medium. By reference to the activity of the language user, speaking and writing are said to be active or productive skills whereas listening and reading are said to be passive or receptive skills. The classification has been advantageous for language teaching because proficiency in a foreign language presupposes that students practise all the four skills (Hirvonen 1971a, 23).

As a communicative activity speaking is usually face-to-face interaction, constituting part of the conversation or dialogue. Thus what

is said depends on what else has been said in the interaction. Consequently, speaking is part of the exchange of words which consists of receiving and producing. In interaction speaking is accompanied by gestures, facial expressions, and as a matter of fact, the utilization of the whole body. These nonlinguistic elements are conveyed visually. Previously, we precluded such phenomena from the instructional aim of foreign languages.

In a communication situation between two people both partners are active because the sender attempts to make himself understood by means of the linguistic symbols and the receiver attempts to understand the message conveyed by them. The speaker says something and the hearer listens. Likewise, the reader actively tries to understand what the writer wrote. Therefore, the distinction between active and passive activities is not justified. In speaking the individual produces by means of his productive skill. In receiving messages he actively uses his receptive skill. In the same way, when writing the person needs her productive skill whereas in receiving the visually perceived symbols she actively attempts to understand the message. This justifies the distinction between the productive skills and the receptive skills.

The activities observed in communication may lead to talking of the four skills of language proficiency. But it remains to be solved whether they are also separate skills in the individual's consciousness. In conversation the linguistic material is orally transmitted by the partners, which means that in listening the sound is perceived as phonetic, and correspondingly, the speaker produces phonetic signals. The phonemes produced and received are unlikely to be very different. Applied to a given language, mastery of the phonetic code would thus be a unified skill. Since the roles of the speaker/listener change with ease, we might assume that encoding and decoding a message takes place with equal ease. Some evidence has been shown of a considerable, but not complete overlapping between the two processes. Leeson enumerates the following factors to prove this connexion:

1. The nature of the linguistic code common to both, positing, for mutual comprehension, immediate access to a broadly similar syntactic and semantic network.
2. Listening and speaking simultaneously is extremely difficult; one or the other activity will be seen to be clearly subordinate on a given occasion.
3. The effects of delayed auditory feedback and other enforced changes of signal reception/articulation relationship.
4. The research in the field of language acquisition has shown clearly that decoding antedates encoding and in some respects exerts a triggering effect upon it.
5. The fact that the speaker's utterance forms at least a part of the stimulus for the hearer's response in any conversational situation. (Leeson 1975, 77—78)

Mastery of the phonetic and orthographic systems of language can be characterized as motor-perceptive skills. In the first language they are abilities that the person has acquired or learnt. The linguistic elements of

the phonetic and ortographic codes are united into larger units, which are then utilized in forming certain structures by means of networks at different levels. This higher-order structure of the code reveals the morphologic, syntactic and semantic factors. Since the different elements of the language usually have their equivalents in the phonetic and ortographic encoding, we might conclude that the communicative media have the grammar, or the object language in common.

On the basis of the preceding discussion we may reject the view of the four separate skills of communicative language proficiency. The different units of the object language are combined and organized according to the rules of the language at the levels of phonology, syntax and semantics. When the linguistic elements are mastered as the combinations of the rule system, we may speak of integrated language skills or integrated language proficiency, which thus would be the common factor for the different communicative media in transmitting verbal messages.

If the above classification of the language skills is accepted, the remaining aspect of the unified language proficiency is the separate means of transmitting messages: auditory and visual media. Since the language knowledge is an individual's accomplishment, the auditory medium will give the term 'oral skills' or 'oral proficiency', because the signals are sent by the mouth and received by the ear. The corresponding ortographic medium will give us the term 'written skills', 'written proficiency'. This distinction will then produce the oral and written modes of communication.

In connexion with oral and written communication, we have spoken of understanding the speaker and of being understood by the others. This individual psychological process will refer to the level of integrated language proficiency. In our instructional model for foreign languages, communicative proficiency was characterized as the skill of processing discourse. The focus is no longer on the object language but has been transferred to the use of the language as a communicative vehicle. At the discourse level the person must make inferences on the basis of the language as well as of the subject-matter, regardless of the communicative mode. This psychological process can be called interpreting. By interpreting is meant the person's internal process which is needed in the communicative language proficiency to understand and produce messages. The only means of gaining insight into this process is through the person's own products, i.e. when she writes or speaks in a communicative situation. This interpreting is bidirectional: interpreting the received message and interpreting in the creation of the person's own product.

We have now distinguished three levels of language proficiency from the lowest to the highest level: the motor-perceptive skills, integrated language proficiency (skills), and communicative proficiency (skill) in a

foreign language. Only the communicative proficiency can be considered an instrumental skill even though corresponding features may be revealed in integrated language proficiency. At all three levels there are two media in terms of communication: oral and written. When a person masters the processing of oral and written discourse in a foreign language, we may say that she has learnt the language, or that she knows the language. This definition does not however express the quality of the knowledge, which is a matter of evaluation.

The above distinction seems fairly clear-cut but the difference between integrated language proficiency and communicative proficiency may cause problems without further clarification. The communicative ability is not always above the integrated language level. A person may be fairly fluent with colloquial expressions in everyday communicative situations but be almost incapable of participating in a discussion at an elaborative language level. Therefore, a more detailed description of the instructional aim must be given, that is, whether communicative proficiency is related to communicative situations, rather than to the linguistic context and contextual situations.

Another problem lies in the use of terms. Though 'proficiency' has been preferred to, and it runs through this study, it is hardly possible to keep up a distinction between it and the 'skills' and the collective 'skill' of the ordinary language use. No reference to the behaviourist view is implied by the use of skills. A concise and appropriate label for oral communicative proficiency is *oracy*, which means an ability to express oneself fluently in speech. The corresponding term *literacy* refers to written communicative proficiency. Both terms are related to the native language. Since the concepts imply the mastery of the linguistic code for communication, they could be handy terms in foreign language pedagogy. Thus oracy/literacy in a foreign language would particularly refer to communicative proficiency.

Finally, the distinction between 'understanding' and 'interpreting' is also difficult to sustain other than for the analysis of the concepts. Thus interpreting may sometimes be replaced by understanding or 'comprehending', particularly in reference to receiving messages.

6.2.2.2. Foreign language skills in terms of neurolinguistics

Neurolinguistics is an interdisciplinary field of study, in which an attempt is made to understand the functions of the brain and the language behaviour. The basic theoretical assumption is that language is physically represented in the brain, so consistent correlations may be made between aspects of a grammar, language use and brain function and brain structure. (Whitaker 1975). The evidence comes from EEG records of brain activity in normal subjects and in neurological patients.

Uniting the data obtained from the brain functions with how language functions, Ellegård (1977a) has outlined a model for the verbal mechanism of the brain. It may serve as an aid in constructing hypotheses of foreign language learning. We must, however, remember that the knowledge of the brain functions is imperfect.

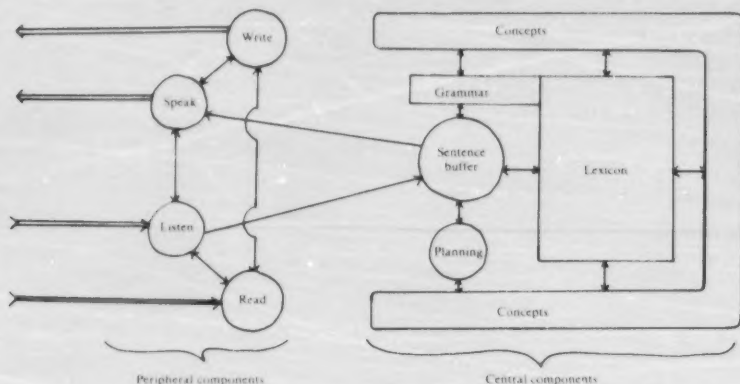


Figure 4. A scheme of the language mechanism in the brain according to Ellegård (1977a).

The model is interpreted as follows: When we want to say something, the impulses go into the vocabulary, from where the appropriate words are chosen. In some cases it is a ready phrase or even a complete sentence, because we use a lot of clichés. In the sentence building centre the words are united into sentences by means of grammatical rules available in the grammar component. The sentence building centre then has a total picture of the sentence that we are going to say. By means of this abstract model the speech centre is capable of building an impulse sequence of the sentence to the speech organs, which then produce the spoken sentence as a physical manifestation. Alternatively, the impulses may proceed to the writing centre, as a rule, probably through the speech centre. In that case the impulses to the speech organs are hindered.

Perception goes in the reverse order. The sound sequence which we hear is analyzed in the hearing centre. The analyzed sequence is stored in the memory, possibly in the sentence building centre, which also in the producing phase acts as a kind of buffer for storing the sentence in the short-term memory. After it the meaning of the sentence is analyzed by the aid of the vocabulary. There the words are identified and connected with their meanings in the form of concepts. The grammar component

may also be involved in the interpretation but can be partly left unnoticed. The interpreted sentence is finally incorporated into our store of experiences by the component which Ellegård calls the planning centre of the organism. What is incorporated is the contents of the sentence rather than its form. Though the memory stores words, phrases and networks between words, every sentence, however, comprises elements unique just to that sentence.

Reading is related to listening almost in the same way as writing is to speaking. Speaking and listening are also interrelated, and so are the writing and reading centres, which is plausible on the basis of the interaction between production and perception. It is very unlikely that writing as a conscious and hardly automatized action is steered by any particular centre. Rather, the question is of a temporary connexion between the speech and hearing centres.

There are other models for foreign language learning which have been developed on the basis of neurological research. Krashen (1977) for example, has created the so-called Monitor Model, in which language *acquisition* is opposed to conscious language *learning*. Language learning involves conscious rule application, the use of the Monitor. Krashen hypothesizes that there is greater Monitor use in a foreign language than in a first language. Related to this model, there is also empirical research on the hypotheses about the hemisphere lateralizations of language in the brain, in different stages of language learning (e.g. Galloway & Krashen 1980).

6.2.2.3. Empirical analysis of language proficiency

The neurolinguistic research shows that language proficiency, at the object language level at least, is fairly unified. This view is supported by empirical findings. Measurements on language proficiency in the different media have shown high intercorrelations which in factorization have formed one broad factor. This can be considered evidence of unitary language proficiency.

To distinguish verbal aptitude from the acquired language knowledge Davies (1971) investigated the first-year secondary school pupils in their skills of French ($N = 1800$). He employed as predictive tests parts of Pimsleur's Language Aptitude Battery and the Modern Language Aptitude Test by Carroll & Sapon and the Moray House Test of verbal reasoning as well as the researchers' own French tests. Factoring produced one broad factor which was interpreted as a factor of intelligence. The multiple regression of the various variables was .70, but since the verbal reasoning test and the aptitude battery alone gave an index of .60, not much can be concluded of the language proficiency. A restriction of the study is that no great language knowledge can be

achieved in one year. The finding, however, supports the view of the common factor in verbal action.

In Sweden the German skills of 134 pupils in the 7th form of secondary school were measured by many different tests, which strongly correlated with each other. A factor analysis gave four separate factors: pronunciation, verbal fluency, aptitude, and a unitary knowledge of vocabulary and grammatical structures — a very essential factor in language proficiency. The factorization did not then support the separate linguistic skills. (Löfgren 1969, 49—50).

In connexion with another investigation, Löfgren refactored the variables. To balance the test battery he dropped some subtests and united some of them. The outcome was similar except that the intelligence which originally was based on two factors formed one single factor of aptitude. The factors were then: (i) vocabulary and grammar, (ii) aptitude, and (iii) pronunciation. (Löfgren 1972a, 33). Löfgren tested this structure of language knowledge with the pupils in the 9th form of the comprehensive school, and considered it feasible (Löfgren 1972a, 37). Evidence from the measurements at different school levels shows such a similarity that it is not merely the question of a single phenomenon at a certain level.

With a new computing programme for comparison of factor structures, Löfgren tested the null hypothesis of the similarity between the structures of the 7-form pupils and the 9-form pupils. The outcome made him reject the null hypothesis (Löfgren 1972b, 7). This result can be considered exceptional, and for computational reasons it may be unreliable. Other findings support the structure of the hypothesis (e.g. Leino 1972, 28).

Leino measured the knowledge of English of students in the second form of the senior secondary school. The different skills were tested by tasks of recognition and production as well as by various integrative tests (reading and listening comprehension, translation, free dialogue production). The oral skill was assessed by the teachers. The first unrotated factor explained 47 % of the variance of the variables, and 74 % of the total variance. The varimax rotation did not give a simple structure, and the author concluded that language proficiency is a fairly unitary structure. (Leino 1972, 18—19, 28). The small number of subjects ($N = 64$) and the case study type of research (one school) may limit generalizations, but the findings are, however, in conformity with other evidence.

Oller & Hinofotis (1980) report results from the data from 159 Iranian university students who took the five subparts of the Test of English as a Foreign Language (ETS) plus a cloze test and a dictation of the TOEFL. The subparts were Listening Comprehension, English Structure, Vocabulary, Reading Ability, and Writing Ability. Once the general factor was partialled out, practically no variance remained to be

explained. In other words, in spite of the fact that there were several tests that required reading comprehension, vocabulary, and structure, no unique factors were needed to account for the variance in those tests. Although we do not know exactly what the tests were like, response modes, multiple-choice questions for listening comprehension, for example, might suggest that the different skills could not actually be singled out. All the tests were focusing more on the linguistic aspects (lexicon, syntax, structure) of language proficiency.

In another study, reported by Oller & Hinofotis (1980), 106 foreign students at Southern Illinois University took the FSI Oral Interview with its five subscales along with a cloze test, and three subparts of the Placement Examination used by the Center for English as a Second Language. As a result of factoring, they concluded that although the existence of a substantial general factor seemed to be demonstrated, the possibility remained that there was some unique variance that was associated with the FSI Oral Interview. A two-factor explanation was supported by a varimax rotated orthogonal solution derived from the principal component analysis. An oblique two-factor solution, however, revealed a .71 correlation between two similarly differentiated factors. Thus the evidence for a speaking factor was not completely convincing but neither could it be ruled out.

In a third study reported by Oller & Hinofotis (1980), 51 of the above-mentioned subjects also took the TOEFL. A varimax rotated solution for the 51 subjects over the 14 tests indicated three orthogonal factors, which the authors tentatively labelled 'reading/graphic', 'oral interview', and 'listening' and which accounted for 39, 38, and 23 percent, respectively, of the total variance in the matrix. Thus the existence of a substantial g factor could be demonstrated, but the possibility of unique variances associated with subtests aimed at separate skills was not ruled out. Criticism against this study can be presented because there were so few subjects in proportion to the number of variables, as a consequence of which the factors obtained are not quite reliable. The factors did not arise on the basis of the separate components of skills: receptive/productive or of the four modes: speaking, listening, reading, writing. This makes one wonder whether the factors were only test-bound so that no generalizations could be made.

Scholz & al. (1980) primarily concentrated on the differentiation of skills, rather than components of skills, such as listening, speaking, reading, and writing, or possibly, some combination of subsets of these skills such as listening/speaking versus reading/writing, or perhaps listening/reading versus speaking/writing. The subjects were 182 students enrolled at the Center for English as a Second Language (Southern Illinois University). The largest language backgrounds represented were Farsi, Arabic, Spanish, and Japanese. Twenty-two test scores were obtained. In factorization, the principal component revealed

that 51.4 % of the total variance could be accounted for by the g factor. All the tests shared a good deal of variance with the general factor, with the exceptions of the Multiple-Choice Listening Comprehension and the Listening Cloze (Multiple-Choice) Format scores, and the FSI Oral Interview Accent Scale. Once the g factor had been extracted, residuals remained relatively small. The authors concluded that the results failed to reveal any unique variance which could be associated with separate language skills or modalities.

The authors also applied a varimax rotation with the solution of four factors. The FSI type Oral Interview, with the exception of the accent rating, loaded heavily on one factor, and the CESL Placement Test (reading) on another. However, all the other experimental tests were scattered over four factors in no discernible pattern which could be associated with the divisible competence hypothesis. The varimax rotation represented only a six percent gain over the single-factor solution.

Although the statistical results are convincing in the previous studies, criticism could be aimed at the tests. They focus on measuring much of the same linguistic mastery, and do not actually distinguish the skills as such, despite their titles. Many tests seemed to concentrate on separate words or short phrases although they were put under the title of listening comprehension or reading, etc. The oral interview evidently measured language proficiency on a broader basis in spite of grammar, vocabulary, fluency, and comprehension being rated. This was perhaps the result of having open-ended responses.

6.2.3. Proficiency in spoken language

In the empirical part we shall concentrate on oral proficiency and particularly investigate its evaluation. Two aspects then are associated with spoken language: speech comprehension (usually called listening comprehension) and speech production (i.e. speaking or oracy). Therefore, a more detailed discussion of the two processes is appropriate here. The next two sections will be devoted to it.

6.2.3.1. Speech perception and listening comprehension

A number of stages can be distinguished in perception which do not necessarily follow in temporal order. They are: (a) auditory, (b) phonetic, (c) phonological, (d) lexical/syntactic/ semantic. In the auditory stage, the signal impinges on the ear as a continuous phenomenon, but since we are familiar with speechlike phenomena, we are able to make some phonetic sense out of it. Once we have recognized the incoming signals as speech, we can break them up into segments

corresponding to chunks, possibly of syllable-like size. In the phonological stage, knowledge of the language code helps to interpret the signals correctly. Some experiments have shown that semantic information is not needed in this phase. The only rule to be followed is that the order of phoneme-like segments do not conflict with the distributional rules of the language. What is at least known about speech perception is that it is not processed phoneme by phoneme. (Cohen 1975).

Miller and Isard (1963) made experiments on people's ability to repeat the sentences which broke against the grammatical, semantic and pragmatic rules of the language. They assumed that it was easier to imitate an intelligible sentence. In a second experiment the intelligibility of the same sentences was measured in a situation disturbed by noise. The experiments showed that the linguistic rules influence our ability to understand and repeat sentences to a large extent. Violations against grammaticality caused most difficulties, and next after that, semantic violations. The researchers concluded that the experiments would give theoretical problems because they did not support the probabilistic view of the transmission of linguistic information. Since linguistic rules are not laws of nature but descriptive ones, there will arise issues for psychologists dedicated to a search for causal relations. (Miller & Isard 1963).

The following conditions must be fulfilled by the listener for speech comprehension to be possible for him: he must have a well-functioning auditory organ; a sensory-motor link which enables him to segment the acoustic continuum as controlled by the mechanism of production; a knowledge of the phonological constraints of the language used; a knowledge of lexical/syntactical/ semantic structures, i.e. a mastery of the linguistic code according to which the message is framed. If these demands are satisfied, understanding will be feasible though listening circumstances can be distorted by noise. Nevertheless, a language user is able to fill in the missing parts on the basis of his familiarity with the language. He projects knowledge of the possible shape of the message onto the incoming signal and thus understands the missing parts. (Cohen 1975).

What is worth noting about segments is so called parallel processing of information, that is, cues for the perception of one particular segment are carried by adjoining segments. This has been shown by experiments in which the natural linkage is severed. The individual items by themselves do not give enough information to the listener even though they are presented at a slower than normal rate. This phenomenon can be generalized over the lexical/syntactical/semantic stage, too: individual words spoken carefully one by one and constituting a sentence will, however, provide difficulty in understanding even though they are presented in the right order. (Cohen 1975).

As for the learner of the foreign language, he will be capable of managing perception stages one and two because he is accustomed to listening to speech in his mother tongue, in spite of the fact that he does not understand the foreign language. His knowledge of certain phonological regularities makes him overhear in the foreign language some features which do not occur in his native language. The listener is then able to mimic the incoming speech to some extent, but in case of longer stretches in the foreign language he will be completely unable to mimic.

For perception stages 3 and 4 it is necessary to have a knowledge of the language code. In spite of the apparently linear proceeding from the left to the right in listening comprehension, we must remember that higher-order functional structures of the language are based on a hierarchical system, which usually differs markedly from the surface structure of the sentence. In the opposite case sentences show great differences in the deep structure. (Leeson 1975, 86).

In this study decoding and encoding processes are theoretically combined into a psychological skill of interpretation because the starting point is in discourse processing. The view is primarily based on the linguistic similarity, and not on the very processes. The processes of perception and production in themselves are different to some extent, but the material processed, i.e. the language, is similar, which justifies the view of the discourse language.

For evaluation of such high-order language proficiency, it is not essential to analyze the physical performance because, in principle, people are physically capable of listening, if they have a normal hearing apparatus. This is not to deny that there are human differences in perceiving speech, even in case of the native language. People differ in their ability to listen to speech, in their power of concentration, and also, in their short-term memory. Cohen (1975) calls such factors together auditory verbal intelligence. For these reasons, it can be stated with much certainty that such discrepancies underlie performance in a foreign language too.

In addition to the auditory verbal intelligence, differences among people are manifested in the ways they understand meanings in producing and receiving speech. In communication, the speaker and the hearer are individuals. Their proficiency in language use is different, which makes the meaning of the message different to each of them, in spite of the fact that it is physically perceived and produced correctly. In addition to language proficiency, many factors contribute to understanding the message: discrepancies in cognition between individuals and disturbances in the communication situation or channel, familiarity with the subject-matter, attitudes, etc. Thus, for example, members of a group who are listening to the same speaker understand or at least conceive of what they hear slightly differently, both in relation to the speaker and to other listeners.

6.2.3.2 Process of speaking and speech production

Different theories and views share the concept of the encoding process as a hierarchy. This means that the superior levels at the hierarchy subsume its inferior levels and that the whole will be greater in some measure than the sum of the parts, since each level has a reciprocal effect on the other levels and exerts an influence on the whole. (Leeson 1975, 32).

Current thinking on encoding is based on the accumulated evidence both from psychology and neurology. Thus the division of the hierarchy of encoding into three units is done on the basis of greater or lesser automaticity. Leeson (1975) gives three levels: a) the strategic, b) the generative, and c) the integrational. The strategic level imposes on the utterance its broad shape, derived from the motivational elements which constitute the 'intention of saying a thing'. The exteriorization of this intention is accomplished by the generative and integrational levels of encoding. The former may be thought of as the level of structural, syntactic decision-making whereas the integrational level implements the decisions taken at the generative level and exteriorizes them in acceptable forms of the natural language in question. Once the speaker has started his utterance, the discrete speech events are run off with varying degrees of automaticity dependent upon the hierarchical level involved: the integrational level with maximum automaticity, the generative level with a high degree of automaticity, yet subject to influence from choice and decision-making processes of a less automatic order. (Leeson 1975, 36).

Since the hearer/speaker of a foreign language is 'able' to hear only sounds included in the sound system of his native language and since he replaces foreign sounds by phonemes of his own language, it is natural that he uses a typical phonetic value of his own language, also when he pronounces a foreign-language word. (Lehtonen 1972, 21). Mishearing foreign language as well as actual mispronunciations due to it, are connected with the differences of the sound systems of the mother tongue and the foreign language. Learning new sounds means learning new distinctions. To replace the paradigm and syntagm of the phonemes of the native language, the student has to acquire a new 'perception filter', which picks out from the speech stream cues relevant to sound units of the foreign language, and categorizes the observed sound units into sound classes of the foreign language. (Lehtonen 1972, 25).

Speech is not, however, so constituted that separate sounds in sequence are articulated. There are many hints that the operation impulses of the speech organs are organized into larger units, whose complex nerve impulses lie as ready programmed models at a deeper level of the brain or the central nervous system. When speaking, a person does not thus regulate orders of production of individual sound

units but triggers action models at different levels. (Lehtonen 1972, 44).

At the functional level the speaker must master such models with automation. Therefore, it cannot be necessary to test production and discrimination of sounds separately when communicative proficiency in foreign language is measured. This view is fortified by the importance of the correct rhythm and intonation to the hearer. According to Lehtonen et al. (1977, 60) the major communicative function of intonation seems to have something to do with the cohesion of discourse: it serves as an anticipatory cue for the hearer about what sort of general structure of utterance is to be expected. This seems to imply that the communicative value of rhythm and intonation is more primary than that of word stress and segmental features of utterance.

6.2.4. Foreign language learning in school instruction

Knowing a foreign language means mastery of the vocabulary and grammar on one hand, and on the other hand, mastery of the practical mechanism for producing and interpreting utterances which employ the vocabulary and the mechanism. The square boxes in the scheme on page 56 represent the development of the vocabulary, phrases and grammatical rules, whereas the circles focus on the mechanisms of production and perception.

In discussion of language learning, the acquisition of the mother tongue and foreign language learning are often compared. The issue is whether these processes are similar or different. Without going deeper into the problem, we may, however, consider the processes different, among other reasons, because the first language is acquired in the form given by the environment. The concepts and meanings are the language games of the community. The games are gradually shaped in the consciousness of the child who uses the language in context and contextual situations.

A foreign language learner already has an established set of concepts which must be replaced by the symbols and meanings of the foreign language. The natural process of learning, similar to the acquisition of the first language, is then too slow and also, more difficult due to the first language. As a consequence, we must accelerate and economize learning in comparison with the natural process. This view must underlie all school instruction of foreign languages.

Ellegård (1977b) suggests two alternatives of accelerating foreign language learning. In employing the natural process, learning must be concentrated and made more efficient by increasing the number of experiences in the foreign language. The other strategy is to systematize the learning process itself. Those who complain of the unnaturalness of school learning, and of the restrictiveness of situations and the language

use in schools, in other words, all the conditions which deviate from the situations of language use in real life, can be said to act in favour of the first alternative. These views have primarily been raised in connexion with the demands for teaching communicative skills in foreign languages. The ideal conditions for learning are considered to be those which resemble the natural circumstances as much as possible. Nevertheless, it is unrealistic to think of creating such opportunities for school instruction. What is more, the very learning process would, however, be slower than in the acquisition of the first language, just because the same amount of time is not available. Also, it is likely to be more difficult due to the earlier acquired concept system. Therefore, as Ellegård (1977b) puts it, the aim for the school strategy cannot be that it is as effective as the natural strategy but rather it must be more effective.

For instruction, it is also important to consider that many factors influence a person's learning ability. One of the brain hemispheres has been observed to be dominant, which causes differences in cognitive learning styles. In addition, e.g. phonetic encoding, sense of grammar, rote memory, personality traits, and the acquired knowledge and skills, strongly contribute to the numerous ways of learning a foreign language (Diller 1976). However, in school instruction it is perhaps too frequently assumed that everybody learns in the same way.

For the above reasons we may advocate the systematic arrangement of the learning processes in order to achieve greater efficiency than in the natural circumstances. This view also underlies the operationalization of the foreign language knowledge into discourse processing, and then into the instructional model. This model is then meant to guide teaching foreign languages in a communicative fashion.

The earlier acquired first language must be utilized in foreign language instruction. As a natural consequence, the systematizing of learning leads to cognitive learning theories. For example, Ausubel's theory, which applies to school learning (Ausubel & Robinson 1969), might provide sources for empirical research. Cognitive theories alone are not sufficient, for the use of linguistic symbols requires quite a lot of automation, which the learning theories must also embrace. As evident, different theories must be utilized and investigated to develop a unitary learning theory of foreign languages.

PART II

7. Evaluation of Foreign Language Proficiency

By evaluation is meant the value assessment of educational elements, processes, and outcomes (Heinonen & Viljanen 1980, 11). Evaluation in schools is part of the realization of the curriculum. Carrying out the curriculum is an educational policy, which has been derived from general goals of society. Through evaluation, data are gathered of how and to what extent the written curriculum is realized in the concrete teaching event. This study does not tend to verify the written curriculum but to develop a model for evaluation in accordance with the previously displayed instructional model. Such a model will serve to guide assessment of proficiency in a foreign language.

Evaluation is in this study reduced to the assessment of the learning outcomes. Philosophical discussion of the meaning and utility of knowledge will be omitted as well as the evaluation of the educational goals of the curriculum. We adhere to the language pedagogical curriculum and to the realization of its aims. The higher-order terms 'assessment' and 'evaluation' are used synonymously whereas the lower-order term 'measurement' refers to the concrete act of assessing proficiency, and accordingly, an instrument or a test is a vehicle for doing it. Testing and measuring will be employed synonymously.

7.1. Analysis of language tests in terms of purpose

The purpose for which tests are used serves as a criterion of distinguishing different tests: there are achievement tests, summative tests, proficiency tests, aptitude tests, diagnostic and formative tests. Through achievement tests an attempt is made to determine what has been learnt of a certain syllabus. A typical example is the school-leaving examination (the matriculation test). Proficiency tests are in a mediating position between aptitude tests and achievement tests (Leimu & Saari 1976, 84). Their task is to measure language knowledge or skills in general without being based on any particular curriculum. For that reason the term 'proficiency' runs through this research. Aptitude tests are used for determining talents and abilities that are needed in learning a foreign language. Diagnostic tests serve to find shortcomings in the language knowledge as well as gaps in the learning.

Bloom et al. (1971) divide evaluative tests into formative and summative in scope. The function of the former is to grade students, judge the effectiveness of the teacher and compare curricula. The

formative tests taken by pupils can hardly reveal the effectiveness of the teacher as such, but indicate what has not been learnt yet. Summative tests actually measure attainment of a broader domain after the termination of a longer period of study. Contrary to summative tests, the formative tests attempt to determine the realization of short-term instructional objectives. As the name implies, formative evaluation intervenes during the formation of the student, when the learning process is not yet completed. It points to areas of needed remedy so that instruction and learning can be adjusted accordingly.

Sometimes the term 'prognostic' is used of the tests, by means of which the student's future performance in a foreign language is predicted. In Finland, the prognostic tests are usually replaced by achievement or summative tests, by which an attempt is made to judge how well the student performs the activities set in the aims of the curriculum.

Evaluation in this study is understood as an assessment of the learning outcomes, and thus focus is naturally centred on achievement. We understand achievement in general terms, similar to an attainment to be reached e.g. by the end of the senior secondary school, which then is assessed in the matriculation examination. Since such a test cannot measure any particular material of a single book, for example, but the language knowledge in terms of the broader curricular aims, our concept of achievement approaches the concept of proficiency, as applied in this study.

7.2. Theoretical views of foreign language evaluation

In foreign language testing the student's performance is measured but the types of tests and examinations often determine the methods and techniques which teachers employ in teaching foreign languages (Mackey 1965, 404). In Finland such a backwash effect was seen in the responses of the teachers ($N = 340$), who considered that the school-leaving examination guided foreign language instruction in the senior secondary school to a larger extent than the official curriculum (Pasanen 1977, 132). As a consequence of changes in the school-leaving examination, some change has also taken place in instruction, though the official curriculum has remained the same. For that reason, the method, in a broad sense, and testing must have a unitary theory or model. The demand is particularly justified for the purpose of assessing whether the students have achieved the level of knowledge and skills required in the curriculum. The achievement tests can thus be regarded as a measurement of the realization of the curriculum.

In language testing Davies (1968, 1) sees three strands: language, learning and evaluation. In the past half-century these have disentangled

three main positions of language teaching and testing. On the basis of the relationship between language and learning, we can distinguish the following positions, slightly modifying Davies: the traditionalist, the structuralist-behaviouristic and the psycholinguistic and sociolinguistic positions.

7.2.1. The traditionalist position

There are advocates who see language as a thing in itself: 'real' language exists in texts, and spoken language is neglected. Literature is emphasized, and little consideration is given to the learner. Learning is studying texts, and teaching is by the grammar-translation method. As a consequence, testing is translation of texts. This slightly exaggerated picture describes the well-known method in which classical languages were taught. It was also adopted to the teaching of modern languages in a modified form. The linguistic theory that underlies this position is the traditionalist one, but no psychological theory is involved to take the learner into account. The language is taught by this method just as if it were a thing, a subject of facts like history or geography. The extracts from literature are usually graded from easy to difficult.

7.2.2. The structuralist-behaviouristic position

Those who do not see any difference between behaviour in general and language behaviour consider language a stimulus-response mechanism. The language itself has no connexion with the environment; it is part of human behaviour. Since language is separated from the environment, meaning is little resorted to. Linguistic analysis proceeds — in principle — from sound to sentence. Language is analyzed into its parts, which are called structures. They manifest themselves in phonology, grammar and semantics. This has led to the approach of taxonomies. Teaching begins with the smallest elements. Thus it proceeds from sounds to sentences.

Learning, it is assumed, takes place by generalization, by analogy on the basis of the learnt response. The linguistic theory of this position is structuralist, the psychological theory is behaviourist. Testing is done by providing drills for completion, gaps for filling and multiple-choice answers for selection. (Davies 1968, 3). It is assumed that the tests measure the learning outcomes of the foreign language knowledge. However, the question arises whether the tests indicate that the student knows the language at a certain level or that he knows how to use the language in the communicative situation. This doubt is due to the emphasis in testing on the discrete elements and not on the integrated language proficiency.

Spolsky (1976) calls this type of approach psychometric-structuralistic. He supposes that by 'psychometric' is expressed the wish to describe an objective and reliable assessment. Spolsky's assumption may be partly true of the original rationale, but now we may perhaps take a more neutral position towards the term and regard it only as a reference to measurement of psychological qualities. Of course, in the operationalizations of such features there remains a gap between the concept and its realization in the 'real' world, which the researchers must be aware of. However, we may critically state that since the object of measurement is narrow in scope and difficult to define, the results can hardly provide a basis for firm inferences of the students' actual language proficiency.

A classic example of the advocates of this position is Lado (1961), whose testing theory states that mastery of the language knowledge is assessed by testing the mastery of its problems. The problems to Lado (1961, 24) are based on the contrast between the student's first language and the foreign language. Those units and structures of the foreign language that do not have counterparts in the student's first language or which have structurally different distributions and meanings cause problems to the learner. Lado's approach has been challenged theoretically (e.g. Upshur 1971) and practically (e.g. Carroll 1961, Perren 1968).

The vast research area of contrastive analysis can be said to continue in the same spirit. Its original strong hypothesis, it is true, has been modified to a weaker one. The contrastive analysis has a vital sequel in the form of error analysis. In Finland contrastive research is done by a group who concentrate on the differences and similarities between Finnish and English (Sajavaara & Lehtonen 1977).

Such comparison may be valuable for teaching and testing, but the research produces a large number of details which cannot be taken into account in instruction and testing. A major criticism against this kind of comparison can be raised, because the linguistically analyzed differences or similarities do not give results on the basis of which difficulty or easiness in learning and teaching could be predicted. Further research in that field must be carried out.

The theoretical background of Lado's approach is that language is a system of habits. Thanks to them the speaker/hearer can concentrate on the meaning to be transmitted or received. This kind of foreign language learning gives a mechanistic impression, and the application of the language knowledge remains remote especially since meaning seems to be neglected. Criticism can be presented against Lado's view. In language use it is not very essential to master one particular element in comparison to the communicating of a message. Also, the fact that the elements of the language and the skills can and must be tested separately (Lado 1961, 28) does not guarantee that the student can communicate

and participate in interaction in the foreign language.

Lado has obtained followers who accept his main principles despite some differences in emphasis. Valette (1967) applies his approach primarily to foreign languages other than English, Harris (1969) aims at parsimony and practicality, and Clark (1972) emphasizes psychometrics. All these approaches can, however, be labelled as structuralist-psychometric.

Harris (1969) pays little attention to the contrastive analysis. Although he discusses analytic tests as well as tests of integrative skills, he devotes himself more to the former. The major part of the book deals with the multiple-choice tests which typically tend to measure different components of language proficiency. Harris is, however, disappointed with the present tests because they do not reflect any unitary linguistic theory throughout the test. He finds this a consequence of the incomplete state of applied linguistics as well as of the lack of understanding of the foreign language learning process. Harris expects the theory to evolve from linguistics whereas the view in this study advocates theory from foreign language pedagogy.

Carroll (1972) classifies the linguistic competence as follows:

1. Knowledge of structure (morphology and syntax)
2. Knowledge of general-usage lexicon (vocabulary and idiomatic phrases)
3. Auditory discrimination (of phonemes, allophones, and suprasegmentals)
4. Oral production (of phonemes, allophones, and suprasegmentals)
5. Reading (in the sense of converting printed symbols to sound, i.e. mastery of word pronunciation and stress patterns)
6. Writing (in the sense of converting sounds to printed symbols, i.e., spelling).

The above list serves as the basis of making discrete-point tests. Nevertheless, Carroll (1972) also wants to test integrated performances in the form of listening, reading, speaking and writing. In these the student has to master the language as a whole, by combining its phonology, structure, and lexicon. According to Carroll, it is primarily the quickness of response that distinguishes the mastery of discrete elements from that of integrated skill. If only one point is tested at a time, more time is allowed for reflection than in a normal communication situation, no matter how rapidly the discrete items are presented. For that reason, Carroll goes on with the aspect list of the linguistic competence:

7. Rate and accuracy of listening comprehension
8. Rate and quality of speaking, as in an interview situation
9. Rate and accuracy of reading comprehension
10. Rate and accuracy of written composition.

Carroll finds it important that tests made for categories 7—10 are used if testing is aimed at total language proficiency. In connexion with the investigations into fluency, it has been found that rate in itself does not indicate fluency because a cared and elaborative speech is not very

quick but there are also pauses. The term 'rate' by Carroll, however, illustrates the fact that performance must take place rapidly enough.

Carroll's notion of integrated language performance is narrower in scope than the preceding description suggests. The impression is confirmed by his examples of excellent testing of listening comprehension in which auditorily presented sentences of increasing length and rapidity were to be matched with one appropriate picture out of four presented (Carroll 1972). The student operated at the level of one sentence and not at the level of communication.

Oller's (1973a) view on integrative language skills is in accordance with that of Carroll. Modifying Harris's (1969) model, Oller presents a scheme in which discrete-point tests are distinguished from those of integrative skills.

Components	Language Skills			
	Listening	Speaking	Reading	Writing
Phonology/ Orthography				
Structure	DISCRETE-POINT TESTS			
Vocabulary				
Rate and General Fluency	INTEGRATIVE SKILLS TESTS			

Figure 5. A scheme of language tests according to Oller (1973a). (cf. Harris 1969, 11).

The views presented by Carroll and Oller can be claimed to lead to testing the object language mainly, because integrating is understood as combining the phonological, morphological and syntactical systems and relations at the level of sentences only. It does not mean the actual communicative proficiency as defined in this study. It is, however, noteworthy that skill or skills are involved due to the requirement of integrativeness. Thus the fact knowledge of the language will be left in

the background. Consequently, it can be said that the integrative skills tests attempt to evaluate in terms of the instructional model although the student still operates at the level of the object language mainly.

In another article Oller (1973b) strongly refutes the analytic approach by appealing to the pragmatic language theory. One of its main principles is that totality is greater than the sum of its parts. Obviously Oller understands integrated language proficiency more broadly than Carroll, because he includes psychological and culture-bound habits and customs in the mastery of such language proficiency (Oller 1973b).

7.2.3. The psycholinguistic and sociolinguistic position

The analytic approach to language testing has also been criticized within psycholinguistics and sociolinguistics, the special branches of theoretical linguistics. The various views in psycholinguistics are based on the concept of linguistic competence. In structural linguistics, language proficiency was defined as a mastery of language habits. Chomsky's concept stresses the linguistic creativity in the language competence in terms of a rule system. The rules enable the speaker/hearer to produce and understand innumerable new sentences. As Spolsky (1976) states it, such a view makes questionable the statistical validity of discrete-point tests.

From the statistical communication theory the linguistic competence has an added characteristic, the view that language is redundant. It is revealed in the fact that native speakers understand the meaning of the message, even though large parts of it are covered by noise. This theory enables the testing of foreign language proficiency in functional terms (Spolsky et al. 1968).

In the theory of pragmatic grammar, the corresponding phenomenon is referred to as grammar of expectancy, by means of which a person is able to predict the missing parts of the message (Oller 1973b). Expectancy involves a complex inference process of probabilities on the basis of various cues at different levels. Such levels of expectancy are at least phonetic, phonotactic, morphemic, syntactic, lexic and stylistic (Enkvist & Kohonen 1976, 9—11).

The psycholinguistic position, too, has been criticized within sociolinguistic research for its omittance of the complexity of language use in various communication situations. The key statements can be crystallized into two claims. A language knowledge comprises an ability to use language in certain circumstances, in which the most decisive thing is the speaker's global skill to manage in the foreign language. It follows that the speaker has to master various variants of the language as well as be able to handle them on different occasions. The language proficiency required means mastery of communicative competence.

The sociolinguistic position accepts testing integrated language proficiency, but also strongly emphasizes the functional dimension. For example, according to Jakobovits (1970) the problem of testing is to describe the manner in which an individual functions at the three levels of meaning: linguistic, implicit and implicative meanings. The linguistic meaning is a lexical (dictionary) meaning. The elliptically derived conceptual event which an utterance represents gives the implicit meaning. It is the result of an inferential process. Utterances as a whole have a relation to several aspects of the speaker: his intention, his psychological state, both intentional and unintentional expressions, etc. These implications give the implicative meaning of an utterance. To exemplify it, "Do you have a match?" should not be understood as a question but a request for fire to light a cigarette. An unintentional clue to geographic origin may be revealed phonologically. (Jakobovits 1970, 155—157).

Jakobovits' classification covers, at least partly, Austin's categories of locutions, illocutions and perlocutions. Its scope is wider than that of the philosopher's analysis, which is based purely on the language. Jakobovits' perspective will perhaps be applied to evaluating the first language better than a foreign language, judged by his detailed analysis of the constituent parts. He does not, however, seem to think so himself.

As a consequence of the investigations of communicative competence, the problem of testing it has also been discussed. Wilkins feels that we cannot really test communicative competence because: "At the moment, ... it is true to say that we do not know how to answer ... questions of spontaneous, fluent, and appropriate speech." (Wilkins 1976, 82). The emphasis on culture and its influence on the language has resulted in peculiarities of testing. An attempt is made to test the culture-bound and sociologically based habits and customs. There are at least two objections to such a procedure. First, one can ask whether a single behaviour can be generalized to all individuals in a given society or culture. Second, it is doubtful whether such facts are really relevant to verbal communication in the foreign language. Seeley (1976), for example, tried to write a multiple-choice item test which would have only one correct answer for all Latinos who answered the item. The item had a stem something like "Where is a Latin business man likely to invite a client to conclude a large and important business deal?" It goes without saying that there is no one solution. Brière (1980) works along the same line of thinking but suggests that the problem will be tackled by taking into consideration the percentages or proportions of occurrences of the phenomena being studied in terms of the number of occasions in which these phenomena could have been elicited. This perspective seems unreasonable for learning as well as testing a foreign language. That is not to deny that the knowledge of such details can be important in social life, but it is unrealistic to think that the foreign language student should

identify himself completely with the foreign form of life.

With the exception of the cultural items, the foregoing discussion shows that language testing has been primarily steered by linguistic analysis with its variations, and by psychology on the other hand. Another point is that the linguistic and psychological views seem to be applicable to testing both the mother tongue and the foreign language. It is justified from the standpoint that the acquisition of the first language does not markedly differ from the learning of a foreign language. As a matter of fact, psycholinguistic and sociolinguistic theories are based on the competence of the first language. This has perhaps led to an unnecessary emphasis on sociological aspects and language varieties for foreign language learning as well. It is natural that such features are analyzed and isolated in linguistics, but it does not follow that they are all essential to a foreign language learner. Previously we have shown that the student does not learn the whole language. The aims can be restricted and defined by means of contexts, which then lightens the student's burden to some extent. For that reason, testing communicative proficiency cannot be extended so as to cover all the features and variants of the language.

As for the learning side of the issue, we may state that despite the similarities of principle in the learning mechanism between the first language and the foreign language, the very learning event and situation are different. The student does not have connexions with the foreign environment in the same way he has during the acquisition process of his mother tongue. The first language and other languages later acquired may be a hindrance or a facilitating factor in learning new languages. Instruction serves to accelerate learning in order to make it more effective and more economical in regard to time. For that reason it is justifiable to separate learning and testing a foreign language from acquiring and testing the first language skills.

7.2.4. Functional model for evaluation of foreign language proficiency

Linguistic and psychological theories do not actually give the right starting point for the measurement of the learning outcomes in foreign languages. The approach must be based on the instructional curriculum of the language. Its practical realization can be assessed e.g. as learning outcomes. In this study the learning goal of communicative proficiency in a foreign language has been described in the instructional curriculum by a model of discourse processing. So this takes the position of the method — broadly understood — just as the linguistic theories, the types of language description and the notions of language learning have steered teaching in the past. Testing must also follow from the same instructional model if the achievement tests do influence teaching. In the

instructional model for discourse processing the method and testing can be combined so that the operationalized curriculum serves as a theory.

According to the instructional model, evaluation is approached deductively, in other words, the testing procedures are derived from the model. To assess the student's communicative proficiency, the test must measure the discourse language, either in oral or written form. The 'discourse language' is such a general concept that it does not focus on any units of the language but defines that evidence of learning requires a transfer skill which generalizes the learning contents. On the basis of the subject-matter, contents and the contexts defined in the syllabus, it should be possible to determine the language content of the measuring instrument.

The processing skill of the language content is a higher-order concept of language proficiency. Processing is an interpreting skill, which manifests itself in the communicative proficiency of the written or oral modes. By this analysis, the written communicative proficiency comprises reading and writing, and the oral communicative proficiency includes listening and talking, as previously stated.

The language skills distinguished in the activities of communication must be combined so that there would arise models for evaluation of oral and written proficiency. The starting point of the development is the processing of foreign language discourse. It has been defined as bidirectional, i.e. as processing the discourse received and to be produced. The interpreting skill has been used as the psychological term of the processing. An individual's interpretations — encoding and decoding — are open to an outsider only through the individual's own products or performance. This is a general rule of the inferences to be made about the phenomena related to the individual's internal events and processes (Heikkurinen 1971, 148). It follows that producing is also the basic requirement for evaluation of the receptive interpretation. In other words, the only way to know how an individual interprets the speech that he listens to or the text that he reads, is through what he writes or says he understands and through the way he expresses his interpretation. This position justifies our advocacy of the requirement that an individual's communicative proficiency must be evaluated on the basis of his own products.

Since in language proficiency the productive skill is preceded by the receptive skill, we can assume that the student's skill of decoding messages in a foreign language at first surpasses his encoding skill. He evidently both encodes and decodes his mother tongue equally well, although he may experience more difficulties in the planning stage of encoding a message than in decoding a message. Correspondingly in the communicative proficiency of a foreign language, the ideal case would be that the student is capable of encoding and decoding messages equally well. On the basis of this ideal balance, the relation of

production to reception could be regarded as the amount of the individual's oral or written communicative proficiency. The reason why the relation is in the order of production/reception is due to our possibility to acquire knowledge of another person's processes. There are many factors which temporarily limit our capacity to encode messages in a communication situation (memory, tiredness, distractions, etc.). The same applies to the written mode, though the restrictive factors may vary. For that reason, in practice, our capacity to decode surpasses our encoding capacity. Since we learn about other people's interpreting through their encoded messages, the oral proficiency necessarily becomes restricted due to the reduced capacity to encode.

Quantitatively, we might obtain the ideal ratio of a unity between production and reception. In practice, however, it is always below the unity. As a result of the discussion, evaluation of the communicative proficiency could theoretically be viewed as two functions:

$$\text{oral proficiency} = f \left(\frac{\text{speaking}}{\text{listening}} \right)$$

$$\text{written proficiency} = f \left(\frac{\text{writing}}{\text{reading}} \right)$$

in which the terms in brackets refer to processing the discourse language.

Two problems arise from this kind of description. The term 'function' may mislead the reader, because it has a certain meaning in mathematics. Here it is only used analogously. The relation between decoding and encoding is seen as a dependent one, so that when the decoding capacity grows, it also influences the encoding capacity. This is in accordance with the mathematical use of the term. The issue lies in the fact that language proficiency can hardly be expressed in exact quantity terms. The problem will be discussed more thoroughly in the next section. Despite this discrepancy, the analogous model can be said to describe the essential factors for evaluation.

According to the model, listening comprehension can only be measured by productive tasks. Combining listening with speaking produces the functional relation for evaluating oral proficiency. Correspondingly, written proficiency is assessed by productive tasks based upon the texts to be read. Speaking and writing can naturally be in isolation from listening and reading, but the contrary is not possible. In the application of the functional model, the subareas of communicative proficiency in terms of media will be evaluated. Such an ideal, but practicable model requires two types of tests only: one for oral proficiency and the other for written proficiency. To exemplify such tests, we can cite various interview-type tests, dialogue, reporting, summary, reply to a newspaper article, answering questions, etc.

Although certain problems can occur in carrying them out, in the rating procedure for example, these may not be insurmountable obstacles.

If by a model an attempt is made to describe components and relations abstracted from reality for understanding a certain phenomenon, the functional model serves to describe oral proficiency as a function of speaking and listening comprehension, and written proficiency as a function of writing and reading. In practice, however, tests according to this model cannot always be carried out due to various restrictions. Then a kind of simulation may be resorted to. If it is to meet the criteria of the functional model, the essential features must be preserved. In that case, the simulation can be regarded as an operationalization of the actual functional model for evaluation of communicative proficiency. The simulation must, however, include the central features: discourse processing and productive tasks. It follows that simulations can primarily be based on changes in the communicative modes of language proficiency. Two simulations can be presented for the functional model as follows:

Discourse processing

Listening comprehension

Reading comprehension

Product

by writing

by speaking

The numerous variety of language tests that we described in the preceding sections may also attempt to evaluate communicative proficiency in a foreign language. Some people might then argue that they also are kinds of simulations of communicative language. They cannot, however, be regarded as simulations because they do not meet the criteria of the essential features of the functional model. Some of these tests are at best operationalizations of the integrated language proficiency but not of communicative proficiency.

We have derived the two simulations above deductively, and their practicality is based on a logical analysis. If we wish to find some empirical content from the simulations, we must make hypotheses about the relationship between writing and speaking. These hypotheses are then tested by performing experiments. The ensuing empirical part of this study works along these lines.

7.2.5. Quantification of language proficiency

The process of quantifying involves two operations: the definition of the characteristics to be measured, and the construction of a measuring instrument. The defining must be made at two levels. First, a descriptive, conceptual definition is needed to indicate in general terms what is to be

measured. It also implies that general procedures may be suitable for measuring the characteristic. Second, an operational definition indicates precisely what was measured by specifying precisely how it was measured. (Ebel 1972, 82).

In this research, communicative proficiency in a foreign language has been instructionally operationalized into discourse processing. Evaluating it requires productive performance. The model for evaluation actually serves as the operational definition of the measurement. The criteria for good definitions cannot be their correspondence with some external reality, since human characteristics are defined by the human mind. Therefore, their true essence cannot be discovered outside the human mind in the real world. (Ebel 1972, 82). For that reason, critical discussion and further research may corroborate or disapprove the criteria for the definitions presented in this study.

The quantity of the student's language proficiency can be described in two ways: the student's achievement is expressed (i) as the proportion of what he learned to what he should have learned, or (ii) as the proportion of his class or peers who learned less than he did. The first way is called criterion-referenced (domain-referenced, universe defined), and the second is known as norm-referenced measurements. (Ebel 1972, 83). Popham (1978) describes the first type: "A criterion-referenced test is used to ascertain an individual's status with respect to a well-defined behavioral domain."

It must be noted, however, that theoretically, we can speak of the quantity of language proficiency but in practice, it is almost impossible to measure the quantity. Many uncontrolled circumstances influence the measuring instrument. One single measurement at its best may give a narrow picture of an individual's language proficiency. The quantity cannot be relevantly categorized. If the measuring instrument contains several tasks, we usually start from the agreed hypothesis that e.g. a student with four tasks correctly answered, knows the language better than a student who solved only two tasks correctly. Great variation is likely to occur between individuals: even those who obtained the same score may have reached it through different uncontrollable factors. Heikkurinen (1971, 131-153), among others, discusses problems related to educational measurement.

Tests of language proficiency are obviously capable of ranking the testees in regard to the hypothesized language proficiency. As we usually speak of measuring instruments and the quantity measured on them, the names will be used here, but we must bear in mind what was said about them above.

In a norm-referenced measurement, an average level of performance for a specific group serves as the middle point of the scale. Individual performances are then distributed above and below that average level

according to the scores obtained. In a criterion-referenced measurement, the scale is between the extremities: the top of the scale indicates perfect mastery of a given ability, and the bottom of the scale indicates complete absence of the ability. Criterion-referenced measurements of achievement indicate what the student knows or can do in absolute terms, in relation to the defined behavioural domain. The scale of good — poor performance is, in practice, based on some criteria determined by the purpose of the test. When the scale of good — poor is based on comparison between performance of the peers, it necessarily becomes relative. Accordingly, these aspects give us the terms of absolute and relative evaluation or assessment.

Absolute assessment is unconditional to the extent that once the criterion of comparison has been set, it will remain unchanged. The aims of the curriculum serve as the criteria for education, and an attempt is made to define them as forms of measurable behaviour. For that reason, the committee for reforming pupil assessment in schools wants to replace the term absolute evaluation by aim-achievement assessment, because the question is of evaluating the achievement of the aims (Komiteanmietintö 1973, 64). In relative assessment, an individual student's performance is viewed in relation to other students. Nevertheless, the procedure involves absoluteness in the sense that the measurement is directed to the aims of the curriculum.

Both views have been advocated and criticized in measuring learning outcomes. Relativeness has been claimed to obscure the idea of what the student knows well or poorly (Ebel 1972, 85). Absoluteness is said to be better adapted to mastery learning (Carroll 1963b, Bloom 1971), individually prescribed instruction and to building a science of education (Ebel 1972, 85). There seems to have occurred some confusion among researchers about the criterion-referenced measurements. An attempt to determine the state of the art is made by Berk (1980). Criterion-referenced tests may, however, be less suitable for many important areas. They do not provide enough information of the achievement of students, because the level of performance, in relation to other students, is ignored. For example, when decisions on selection have to be made, they are inadequate for it. On the whole, we can, however, note that the difference between the two test types, is perhaps not so great, but rather the question is of the different purposes of the tests.

The analytic and taxonomy-based way of evaluating language proficiency enables absolute or aim-achievement assessment. Since the analysis is based on the linguistic elements, we cannot be sure of the mastery of the language as the network of relations. The complexity of relations in integrated language hinders an exact analysis. Along the dimension of mastery— nonmastery it is not possible to determine precisely what level or constituent part is concerned. Discourse

processing is hierarchically an even more complex functioning, which is influenced by many external and psychological factors. The simultaneous network of relations at the hierarchy cannot possibly be specified exactly. Since the learning aims of the discourse language cannot be determined in a definite taxonomy, they do not lend themselves easily to absolute evaluation. As noted in connexion with the instructional model, we must resort to a more general characterization of the learning aims in terms of register, subject-matter, speech situations, etc., when designating the linguistic aims of learning.

In communicative language proficiency, we must also consider the skill side along the dimension of absolute — relative. If it were possible to define the mastery of the skill with that of the native speakers through the principle of dichotomy, absolute evaluation would be feasible. However, the skill is a complex whole manifest at different levels. It is revealed in the mastery of the object language, and it varies in regard to the individual and community, simultaneously. In countless different regards, the level of the mastery of the language may vary with different individuals. It does not, however, prevent people from managing verbal communication as equal native speakers of the language. In other words, the language consisting of idiolects, allows some variation within the norms before the individual's language is considered deviant from the norms of the language community. The intensity of variation is somewhat reduced by the psychological fact that the hearer adapts what he hears to his own norms. Only marked deviances are perceived. Since this is true of the native speakers, absolute evaluation cannot be justified for the assessment of communicative skill of a foreign language, either.

When measuring communicative proficiency in a foreign language, the native speakers might serve as the reference group to determine the average level of performance. However, it is difficult to defend this view because the different registers, dialects, educational level, experience, etc. among the speakers of the same language vary to a greater extent than it is necessary to learn in a foreign language. The curriculum determines what kind of language proficiency is aimed at. Thus a corresponding population of the native speakers can be used as the reference group. Such investigations have hardly been made use of in deciding the level of performance. An exception is a study by Darnell (1968), who measured English language proficiency of foreign students through the procedure of clozentropy. In rating their proficiency, he employed the performance of the native speakers as the criteria of acceptance. Such reference groups might provide a kind of absolute criterium for determining the level of communicative proficiency in a foreign language.

There occurs some disagreement about employing native speakers as the norm. For example, Oakeshott-Taylor (1977) considers it absurd to set up native speaker proficiency as a norm because the native speakers

vary enormously in the scores of comprehension tests. Hendrickx (1977) disagrees with Oakeshott-Taylor and defends comparisons with native speakers' scores as a valuable procedure of evaluating the validity of a test.

7.2.6. Empirical studies of oral proficiency

As we now have a model for assessing communicative proficiency, it can serve as a perspective, from which the previous empirical tests may be viewed and evaluated. Naturally, the testing has been based on other theoretical grounds. Since the empirical part of this study will be confined to testing oral proficiency, we shall only touch upon listening comprehension and speaking tests. When the functional model for evaluation is applied to measuring communicative proficiency, listening comprehension and speech production are simultaneously measured. Very few such tests have been used, and most often the skills have been tested separately. For that reason, the two processes have been distinguished in the following discussion.

7.2.6.1. Previous listening comprehension tests

The earliest listening tests concentrated on investigating the discriminatory ability of sounds and intonation patterns on the basis of Lado's analysis (1961) of the language skills. The English Language Battery (ELBA) developed by Ingram (1968) was similar in structure. Rivers (1968, 290) follows the same tradition in claiming that the teacher should analyze the skill to be measured in order to be able to construct items for the measurement of the elements identified, separately. Also in Finland, Hirvonen (1971b, 48) defends discrimination tests for the matriculation examination with the justification that the preceding instruction is positively influenced if all the skills are emphasized in the final examination. This reasoning is not, however, sufficient for including discrimination tests in the examination, because language proficiency in terms of discourse processing presupposes lower-order skills whereby it is unnecessary to control them separately.

The view is supported by results from discrimination tests. They have proved to be unreliable, obviously for the reason that they measure differences, irrelevant to the higher level of foreign language proficiency. Such low correlations have been obtained by e.g. Hirvonen (1971b, 31), Ingram's ELBA, Leino (1972, 25). Such unreliability cannot be generalized to all listening tests. Once the aspect of understanding is included, reliability will improve. The interaction of the different skills of English was measured by Hirvonen (1971b, 8) through the interpretation of short oral messages and dialogues. The test consisted of

sentences in isolation followed by one question. From the four Finnish alternatives for interpretation which were given on the question sheet, the student had to pick out the one he thought correct on the basis of what he had heard on the tape. The reliability estimate of the test was .80. (Hirvonen 1971b, 31). The Finnish alternatives were transformed into English ones in another version, and the reliability increased to .83. A corresponding subtest of Ingram's battery showed a reliability of .87 (Hirvonen 1971b, 33).

The often employed and cited test batteries are primarily built on the analytical view of the language and tend to combine several subtests into a totality of language proficiency. Such are e.g. ELBA, and TOEFL, which is an American test battery of English for foreigners. Integrative skills tests provide a means of testing language proficiency at a higher level than those quoted above. Hirvonen attempted to measure English proficiency by testing the understanding of spoken language. His main problem in constructing items was to test the meanings of structures (tenses, conditionals, passive, etc.), some clauses, auxiliaries, pronoun references, and understanding certain essential idioms (Hirvonen 1971b, 8). As the examples show, focus was only on the mastery of the object language in the different modes. Integrated language proficiency in a broad sense was not taken into account.

Spolsky et al. (1968) has experimented with listening tests shadowed by artificial noise. This is based on the theoretical view that native speakers can fill in the noisy parts by making use of redundancy so that no break-down in communication is caused. Foreigners who have achieved a fairly good command of the language do not reach such a level of performance. The experimenters regarded the tests as good instruments.

Criticism against them by Johansson (1973) is based on his many experiments. He maintains that failure in them does not necessarily show weak language proficiency. If a student is successful in such a test, he usually has an excellent command of the language, but even a failure can succeed in other situations which are more realistic to language use. Johansson claims that such noise tests give evidence of a special skill demanded of e.g. radio operators, interpreters, and secretaries, under varying conditions of distortion.

Johansson (1973) recommends the replacement of artificial noise tests by dictation tests as better vehicles of measuring overall language proficiency. In his experiments a dictation test correlated with a multiple-choice listening comprehension test at .83. Similar results have also been obtained elsewhere: dictation correlated with ESLPE of overall language proficiency for university students at .86 (Oller 1971), with listening comprehension at .69 (Irvine et al. 1974), and with an integrative cloze test at .82 (Oller & Conrad 1971).

Lado takes a negative view to dictation, reasoning that the tester gives

the hearer the ready word order (1961, 34). Against this it can, however, be claimed that the hearer must analyze speech at hierarchically different levels in order to understand the linearly proceeding speech. Segmenting is not ready-given for this. Harris (1969, 5) again, considers dictation as an uneconomical and unprecise means of testing.

Johansson (1973) states that the dictation technique is improved as a measuring instrument when students are asked to fill in the missing words in the written version of the tape. This, however, can be argued against by remarking that the blank-filling is not then dependent on understanding speech. Moreover, word order and syntactical aspects are lost if the blanks do not form a stretch longer than a couple of words. On the whole, we can remark that dictation has not been demanded for listening comprehension so much as for measuring overall integrated language proficiency. In some countries, it has established itself as a form of testing foreign language proficiency.

As previously noted, there are many factors which affect scores in listening comprehension tests, such as long-term memory or interpretative intelligence. Some researchers believe that these nonlinguistic factors can be eliminated if we do not test the comprehension of a text in its totality but the perception of a much shorter fragment of a spoken text. Cohen (1975) for example, works on the hypothesis that listening skill can be tapped by a test involving the recognition of short excised segments of the spoken continuum, of one or two seconds' duration. He calls this intelligibility and defines it as recognizing the acoustic speech signal in terms of potentially meaningful linguistic forms. Such a recognition is a necessary, though not sufficient condition for listening, but understanding a message presupposes its intelligibility.

All the means of measuring listening comprehension described above are, however, narrow from the communicative perspective of language proficiency. This study has approached the measurement of listening comprehension from the viewpoint of discourse processing. It means proficiency at a higher level than that of integrative skills if it is not confined to concrete situations. Attention is focused on understanding the subject-matter mediated by the language. The language serves as a vehicle of transmitting the message.

Such a view can be seen in listening comprehension tests prepared by the Board of the Matriculation Examination in Finland. Some questions, it is true, have sometimes been formed in such a way that knowing a particular word has been decisive for the correct response. Then testing has been limited to the object language. Lusensky (1972) also suggested that attention should be focused on the informative points of the extract listened to when testing listening comprehension.

Talking of listening comprehension tests may give the illusory picture that they are similar, but they do differ from each other due to

the various forms of questions, for example. Multiple-choice questions can be directed to measuring the understanding of the subject-matter, a particular structure, or even a single word. In Finland, the listening comprehension test has an established form due to its inclusion in the school-leaving examination. Since Spring 1977 it has been a compulsory subtest of foreign language proficiency in the matriculation examination for the senior secondary schools. Experiments for the development of its form and type have been carried out since 1970. At first it was optional, later an alternative with the so-called traditional test type until it was finally made compulsory. Since it also serves as the basis of the experiment of this study, a brief outline is called for.

The procedure is as follows. Students hear a continuous extract of spoken speech on the tape. It is replayed in short sections so that after each section they can answer questions (usually 1—3) given in a booklet, by selecting from four alternatives in the foreign language the one that has the closest correspondence to the spoken text. Before the extract is replayed, the instructing voice on tape asks the students to look at the corresponding questions in the booklet. A period of ten seconds is reserved for each question. Then the section corresponding to the questions is relistened to. Students are allowed 30 seconds to respond to each question, and the total time is usually about 45 minutes.

As a rule, the material used in listening comprehension tests can vary along the continuum: literary material read aloud — spontaneous speech. If a test is built on literary material, it measures listening comprehension only in a restricted sense. (Lusensky 1972, 4). Lusensky obviously believes that listening to a lecture, for example, demands a different skill than listening to a dialogue. The issue does not perhaps lie so much in the listening ability as in the knowledge of varieties of the language code, and their representativeness in a listening comprehension test.

In experiments with English in the matriculation examination, the material has varied along the above continuum. The now established form consists of three parts with different topics in each. Two of them are dialogues or interviews, and the third is a factual story or theme read by one speaker. Geographical variations of the accent are observed in accordance with the curricular aims. So, besides the British accent, American and sometimes Australian readers have been employed. Socially varying accents have also occurred in some dialogues, but on the whole, only so-called educated speech is preferred.

Both from theory and from practice, criticism can be targeted against this type of listening comprehension test. Since questions and response alternatives are given in printed form, reading comprehension is also involved in the test. As for the total battery of the matriculation examination, written proficiency becomes emphasized more than the subtests will suggest. Moreover, stress is laid on the receptive skills of

oral and written proficiency. In terms of the functional model for evaluation, the set of tests do not give a clearcut picture of whether oral or written proficiency is measured and to what extent.

We might assume that success in listening comprehension is facilitated if the subject-matter is handled in the written questions, too. The multiple-choice procedure reduces the respondent's possibilities to interpret the spoken speech. There are two kinds of consequences to be expected: it may be easier for the student to choose among a fixed number of alternatives, or it is difficult, because he may not understand the written alternatives despite the comprehension of the spoken text. Furthermore, distinctions between the alternatives are perhaps based on nuances of the written language which are completely irrelevant to the text listened to. This phenomenon is revealed in the complaints of the alternatives being too close to each other in meaning.

The use of multiple-choice questions reduces the respondent's possibilities. Neither can we be certain of whether the student really understood the heard text, because he does not give his own interpretation but selects the most probable of the test planner's interpretations. A chance of guessing is not excluded, either. As a consequence of these flaws, there occurs casual variability, and differences in language proficiency cannot be systematically registered.

Teachers have also aimed criticism at the chance of guessing, and at alternatives being too similar in meaning. The probability of making equal alternatives is not easy, because appropriate alternatives are laborious to find without resorting to lexical grounds. The respondent may choose the alternative that differs from the others or that corresponds to a situation in reality, without his actually having understood the speech. (Lahti 1977). Lahti's views are based upon his experiments. The results can be considered symptomatic rather than conclusive, because they were not done in properly controlled circumstances.

7.2.6.2. Previous speaking tests

Many researchers and teachers consider it important to test speech but they disagree as to how far particular speech skills can be isolated and validly tested separately from those skills normally associated with reading and writing. Disagreements are probably due to the categorization of the four skills. The skills overlap and are integrated and are not specific to any one aspect. (Perren 1968). Speech testing can be approached analytically or globally. Rivers speaks for the analytic view and opposes an interview test, for example, with the reasoning that speech comprehension as well as other personal traits are involved, too. Rivers defends analytic testing, mainly with objectivity in view and

believes that global proficiency can be assessed by testing separate skills. (Rivers 1968, 296—297).

Objective tests can be considered unsatisfactory as measuring instruments because of their invalidity, unreliability, difficulties in scoring, etc. (Perren 1968). However, the demand for objectivity and the analytic position have guided the construction of speaking tests. Too often it has been sufficient proof of oral proficiency if the student has succeeded in a pronunciation test, out of meaningful context, or in an intonation test of one sentence long at the most. The criteria for spoken language cannot be derived from the abstract concept of language but from concrete spoken language. (Perren 1968).

In the testing of oral proficiency the ultimate aim is to determine whether the student masters the language use in real-life settings. Possible real-life contexts would be such activities as making a speech, dictating onto a tape recorder, talking on the telephone or face-to-face conversation. In measurement a dichotomy might be used: direct and indirect testing. In direct testing an attempt is made to duplicate the setting and operation of real-life situations. Examples of such tests are conversational situations in which the examinee has to converse with one or more people. Indirect tests do not require real-life situations, but are quasi-realistic on the student's part. An example is a test in which the student is asked to describe printed pictures aloud. (Clark 1975).

The direct — indirect dimension is based on the face validity, and Clark seems to think of it primarily as a logic validity and as the opposite to the validity based on correlations. For that reason, he considers invalid such activities as repeating heard utterances, describing pictures or series of pictures, or reading aloud printed texts. Though there is no denying the importance of face validity, it can hardly be considered the primary criterium for preparing a measuring instrument. Indirect testing can provide us with reliable data more practically and economically. Direct measurements limit possibilities to administer tests, mechanical (technical) apparatus are not practical in the assessment of speaking ability in a conversation situation, which calls for observers to score the performances. Thus unreliability due to scoring is involved, however, and weakens the objectivity.

Indirect testing offers potentially more numerous and practical alternatives than direct testing. The appropriateness of these tests must be based on their reliability and empirical validity in regard to direct measurements. The face validity is important on the part of the examinees. If the test seems natural and relevant in terms of real language use, it can be regarded as seemingly valid. This face validity might be studied empirically, e.g. measuring the attitudes of the teachers and students.

Tests of oral proficiency with the emphasis on the speaking ability has been studied in various versions. In the following they have been

classified into two main categories in terms of openness of the testing situations. Another criterion is administrability in regard to the number of testees.

Tests in an open situation

1. a free conversation test
2. a guided conversation or interview-type test
3. a free or restricted speaking test

Situationally homogeneous tests which suit mass testing

4. (i) answering questions
(ii) situation-based short reactions
5. tests of reading aloud for the object language
 - (i) pronunciation
 - (ii) intonation
 - (iii) repetition tests

Users of various tests usually advocate their own tests. However, there are weaknesses and difficulties in all of them. The above classification serves as the basis for the following discussion of speaking tests.

1. Folland (1976) wants to test overall proficiency in a conversation situation and thus rejects most other test versions. To him a group conversation offers a more precise, less subjective and more realistic way of assessing a student's overall language proficiency. The stimulus of the conversation comes from a prerecorded tape which presents a topic and a wide range of discussion points. The testees may take notes during the listening. The aim is to give a maximum aid of ideas so that they can better show their language knowledge. The testees develop the discussion themselves and the examiner does not guide it. The classification of mistakes is an essential part of the test: major errors concern comprehension, very few of which are allowed; several minor errors which do not hinder communication may be allowed for the attainment of a good mark.

2. In a guided conversation the examinee is asked a number of preplanned questions or the situation is the format of an interview. The interview has been successfully employed as a language test for the staff at the Foreign Service Institute in the U.S.A. (Wilds 1975). After the simple social formulae (introductions, comments on the weather, etc.) the interviewee may tell about himself, his family, his job. Then he is led into natural conversation on autobiographical and professional topics. The interviewer tries to elicit the grammatical features that need to be checked. A third element of the speaking test is an optional one: it involves instructions or messages given in English to be conveyed to the native speaker. The following checklist of performance factors is used

for rating the examinee, with the weighted scoring system. The weights are given in the right column.

1. ACCENT	foreign	— — — — —	native	0
2. GRAMMAR	inaccurate	— — — — —	accurate	3
3. VOCABULARY	inadequate	— — — — —	adequate	2
4. FLUENCY	uneven	— — — — —	even	1
5. COMPREHENSION	incomplete	— — — — —	complete	2

(Wilds 1975)

The raters are professionals, which guarantees conformity, even across the different languages, in terms of the performance profiles. A striking feature in the rating scale is that grammar, vocabulary and understanding are strongly stressed. The question arises of how much is gained in regard to indirect measurements, if the accent is ignored and fluency is weighted with one.

3. In a free speaking test the student may either speak freely for a few minutes about a topic he has chosen himself, or there are some restrictions imposed. Free speaking on a chosen topic enables rote learning in preparing for a testing situation. Since topics vary in difficulty, the testing situation becomes different to different students. (Kostera 1979). At the Tampere Language Institute, Finland, they have tried out a version in which three topics are available for testing, such as education, social welfare, sports, religion, the roles of the sexes, Finnish foreign policy, etc. A number of cards are prepared each listing three of these subjects. The student takes a card at random, and is allowed 15-minute preparation time, after which he gives a 3-5 minute talk on one of the three subjects. After this the student discusses the subject with two native speakers. In the assessment of the test, attention is paid to pronunciation and fluency, command of structure and variety used, vocabulary, idioms and style, and comprehension of other conversationalists. (Bacon & Ojanen 1976). No statistical data are reported so that reliability cannot be known. A striking feature about the test is that consideration is given to speaking factors more than in the above cited test type. This application can be placed somewhere between test type two and three.

Tests in which examinees are requested to describe pictures or cartoons can be considered a restricted form of free speaking. Hirvonen (1973) experimented with such a test of oral expression in which the student studied a cartoon for one and a half minutes, and then described it freely. It was not a very successful type of testing because some students had hardly enough to say for half a minute, whereas others aimed at performances longer than the allowed time. The differences seemed to be independent of the language use and its fluency (Hirvonen 1973, 17). Though without cartoons, Stendahl observed similar problems in testing first-year university students in Sweden. Responses

to questions like "What would you show a tourist...?" seemed to reflect the students' imagination and inventiveness in addition to language proficiency. (Stendahl 1970, 5).

The flaws in the preceding test types can be summarized in two wide problem areas. Firstly, the possibility of administering the tests is restricted because the testees perform either alone or in small groups. Besides, the groups have been relatively homogeneous and have represented people in a specific branch. The other problem is associated with rating. Experienced and specialized staff is needed for scoring. As a consequence, administering such tests for general language proficiency can only be realized to a limited extent.

4. With his test of oral expression, Hirvonen aimed at homogenizing the testing situation for a large group. The test consisted of direct questions and of the descriptions of everyday speech situations to which the students were expected to answer with a couple of sentences. There were such questions as "How do you like to spend your free time? Why are there so many car accidents in Finland? What is a leap year?" (Hirvonen 1973, 15). A three-scale rating was applied with the following criteria in terms of communication:

- 0) no answer or an answer that is quite inappropriate to the question or situation
- 1) a comprehensible answer that gives the information asked for or meets the minimum requirements of the situation despite linguistic clumsiness and errors
- 2) a comprehensible, adequate and socially appropriate answer that is idiomatic and fluent. (Hirvonen 1973, 20)

The reliability estimate of the test was .82 and with another native speaker scorer .84 (Hirvonen 1973, 21). Criticism can be given against the narrow scope of oral expression, because the knowledge required in responses was minimal in comparison to that of the school-leaving examination.

Based on Hirvonen's test of oral expression, a working group set by the Ministry of Education (Klemmt & Kostera & Lohfert & Piippo: Sprachlaborprogramme für den Deutschunterricht auf Hochschulebene) suggests a guided speaking test of six and a half minutes, which consists of three subtests: (i) 10 answers to questions on tape, (ii) 10 answers to questions on tape about a series of five pictures (two questions per picture), and (iii) five free reactions on the basis of the speech situations as instructed on tape. (Kostera 1979). The picture series has proved to work well in practice, which perhaps is due to the fact that the students are asked questions about the pictures, and are not allowed to describe them freely.

Upshur (1971) has investigated a slightly different restricted speaking test. He defines productive communication as follows: "as a result of some action by a speaker (or writer) his audience creates a *new* concept." Accordingly, he measures speaking proficiency by the testee speaking, and the examiner listening, forming a concept and indicating what that concept is. If the two agree, the testee has succeeded in communicating.

In practice, the measurement takes place with the testee being shown four pictures which differ significantly on one or two conceptual dimensions, e.g. a person performing four different actions. The examinee must describe with a single sentence, to a visually remote audience, one picture which is randomly selected from the set. The examiner makes a guess, and if they both mean the same picture, the communication has succeeded. A scorer reliability of .99 has been found, and the test-retest reliability for a 36-item experimental test was .92. The empirical validity has been studied with the help of other language tests. (Upshur 1971).

Though the measuring instrument is reliable, it seems mechanistic and comparable to the multiple-choice tests, with the distinction that the selection is done by the examiner. The testee is given the chance to produce a creative and original response, but he is not awarded for its quality, because the only criterion is selecting the right picture. A weakness lies in the fact that even a poor explanation communicates if the examiner is inventive enough. In other words, there is no guarantee of generalizing to the student's communicative language proficiency.

5. The tests of categories 1—4 represent communicative testing, and thus measure communicative language proficiency. Tests of category 5, which also have been employed, measure a subarea of the speaking skill, focusing on the phonological part of the object language. Hirvonen, for example, also gave a reading aloud test. He used isolated sentences in which one sound was rated, and isolated dialogues in which either stress, rhythm or intonation was evaluated. (Hirvonen 1973, 17). The pronunciation test was scored dichotomously, and correlated with the aforementioned test of oral expression. The index of the correlation ($N = 68$) was .53 with scorer A, and .43 with scorer B, and the reliability was estimated at .76 — .78. The interscorer correlation in the pronunciation test was .81, and in the oral expression test .85. Since the correlation between the two tests was low when computed between the ratings of the same scorer, Hirvonen concluded that the different aspects of language proficiency were really assessed, as it was meant. (Hirvonen 1973, 21).

As previously noted, the hierarchy of encoding raises the question of the necessity to measure lower-order skills, because in any case, they are a necessary condition for the higher-order skills at an operational level. Moreover, reading sentences aloud out of context, requires entering into the spirit of the situation, an ability to express the feeling in one's own performance. Such an ability need not be part of language proficiency. Everybody is not capable of playing roles though he may be able to acquire communicative proficiency in a foreign language.

Speaking skill has also been predicted by instruments in which the students do not have to speak at all. The rationale is based on the high correlations with other tests of language proficiency. It has been found

that a listening test together with tests of vocabulary and grammar predict oral language proficiency very well. (Löfgren 1972a, 41). As previously stated, the view on unified language proficiency lies behind the procedure. How well it is possible to predict language proficiency with tests of separate skills obviously depends on the level of proficiency aimed at. For integrated language proficiency, tests of vocabulary and grammar are good predictors. If discourse processing is included, all information is not gained through integrative skills tests. Such a hint is given by Löfgren's regression analysis. The multiple correlation of vocabulary and grammar tests with the speaking tests was lower in predicting free speech than in predicting simpler, more stereotyped answers. (Löfgren 1972a, 41—43). The outcome must be taken with caution because the subjects were comprehensive school pupils whose discourse processing is not at a very elaborative level.

PART III

8. The Experiment

8.1. Defining the purpose of the experimental study

In the preceding two parts we have carried out a conceptual analysis of communicative foreign language proficiency and arrived at its definition, which then was used to sketch an instructional model for communicative foreign language proficiency. The model was utilized to analyze language proficiency and skills in terms of evaluation, as a consequence of which a functional model was offered to appraise communicative foreign language proficiency. All the preceding analyses serve as a theoretical framework for the field experiment of this third part.

The main question after this theoretical development is to put theory into practice, i.e. to try out a measuring instrument based on the functional model. In Humean view, such a deduced model cannot, however, be verified by empiricism. If we accept this view, it is necessary to find a theory built upon empirical findings, which is then tested in the practical application. On the other hand, the Humean statement can perhaps be interpreted as referring to the formal deduction of a model, i.e. the correctness of the deduction cannot be verified or falsified by empiricism but by showing errors against logic in deducing the model. What interests an educationalist is whether the contents of a theory or a model correspond to reality and whether the model or the theory is applicable. In that sense, theory is put into practice.

Research on teaching can be different from basic research, e.g. in psychology or education, in the sense that the former is always connected with influencing people through some kind of activity. This is the reason why pedagogical research can be argued to be finalistic or teleological in character, because the aims serve as a theory for the intentions to achieve them. In this sense, such research is a practical affair and often arises from practical problems.

For this study, it is very unlikely that an oral test will be administered in the matriculation examination, due to its cost and complexity and difficulties in carrying it out. This fact guided the present empirical research to search for other solutions to the problem of evaluation. Theoretical support for the empirical design was gained from a hypothesis, which has been presented on the basis of some empirical evidence. According to the hypothesis, the structure of language proficiency is very much of a unitary factor (like Spearman's g-factor) with no specific factors, or at least none of practical value, to be found after it has been singled out.

Thus, the hypothesis provides the theory to be tested in the experiment. Its advantage is that it has been induced from empirical findings, and not deduced from some other theory, so that no logical error will be made in testing its validity. In spite of the theory to be tested, there is no denying the value of the preceding theoretical development. Instead, it provides the framework for evaluation of communicative foreign language proficiency, for preparation of the measuring instrument and its administration. As a consequence, the experimental part gathers all the threads of this research enterprise, in which theory and practice as well as theory and phenomena are combined into a smooth ensemble.

The empirical study is based on a field experiment. To rigorous scientific methodologists, an experiment means a scientifically strict control over the experiment performed under laboratory conditions. For educational and pedagogical research, it is, however, vitally important to work in the natural environment of schools, so that the actual prevailing conditions can be discovered and can have their effect on the results. When an experiment is designed in such circumstances, complete control over intervening factors cannot be obtained, but that is not to say that a field experiment could not be sufficiently reliable for inferences to be made on the basis of it. The utilization of the experimental design may justify the name of an experiment, although it is made in the field, and not in the laboratory.

The name *quasi-experiment* might be used of an experiment in the field circumstances. Cook & Campbell (1979) discuss quasi-experiments in detail, and define them as "experiments that have treatments, outcome measures, and experimental units, but do not use random assignment to create the comparisons from which treatment-caused change is inferred" (Cook & Campbell 1979, 6). The essential feature of the *quasi-experiment* is thus nonrandomness, and then nonequivalent groups. Since random sampling has been used in this study, we shall maintain the name 'experiment', although it is performed under the field circumstances.

Problems within the framework of the whole evaluation model are too many to be solved in one study, so the object of the experiment had to be limited. Attention was then focused on oral proficiency. There are many reasons for that. For the time being, emphasis is laid on oral skills of a foreign language by researchers and people in general. On the other hand, criticism against listening comprehension tests with multiple-choice questions (e.g. Lahti 1977), and demands for speaking tests in the matriculation examination (e.g. Hirvonen 1973, Kostera 1979), inspired the investigation of oral proficiency.

Further, the empirical study is confined to investigating a test type of communicative foreign language proficiency in which answers to questions are either spoken or written. Both these versions can be

considered operationalizations of the functional model for evaluation. Thus, attention is paid to the measurement of communicative foreign language proficiency as well as to the measuring instrument as a vehicle.

English was chosen as the language for the experiment, because 96.7 percent of the students in the senior secondary schools in Finland study it as their first foreign language, so that proficiency in it is supposed to be at an advanced level. Also, this author's practical experience of teaching English at this level, coupled with his thorough knowledge of the procedure of the matriculation examination in English contributed to the choice. Nevertheless, the language is not thought to be decisive for generalizing the results. Instead, advancement in a foreign language may have its influence on generalizations to be made on the basis of the experiment, since only the students of the third form, i.e. of the highest class of the senior secondary schools, were chosen as subjects. In this phase, students can be expected to approach the level required in the learning aims of the curriculum, which all senior secondary schools in Finland share.

8.2. Problems and hypotheses of the experiment

The main problem of the experiment is whether the structure of proficiency in a foreign language is a unitary factor, with no specific factors, or at least none of practical value. The hypothesis of a unitary factor of foreign language proficiency is based on some empirical evidence, as presented above. To test this hypothesis, an experiment was designed to measure communicative proficiency in a foreign language. Instead of having a number of tests to measure different skills and modes of linguistic proficiency, and then having them factor-analyzed in order to verify a unitary factor positively, this experiment was planned to falsify the hypothesis. This means that the distinction between oral and written responses, when they are assessed on the basis of temporarily restricted reactions to questions heard on tape, would be tested against the statistical null hypothesis of no difference between the answering modes. Thus, this experiment deviates from previous studies in two ways: first, it is an approach to falsify the hypothesis, and second, an attempt is made to test the hypothesis at the level of communicative foreign language proficiency and not only at the level of integrative skills and integrated language.

To control individual differences between experimental units (i.e. subjects), it was thought necessary to have repeated measures on the same elements, which meant, that the same pupils should perform on oral and written modes. Since it was not possible for the same pupils to give both oral and written responses to the same questions, answers used comparing the modes would be different. Differences in difficulty

between the first and second parts, for example, might cause a difference, not due to the effects of the treatments. Consequently, students had to be divided into two groups: Group A, first answering by writing, then orally, and Group B, first answering orally, then by writing. In this way a comparison was made possible between the modes of the first and second parts of the test, respectively.

The students could have been assigned to groups at random to guarantee comparability, but differences in verbal ability were considered to be such an important factor that it might make treatment effects spurious if not controlled. This might especially be the case if classes, small in number, should be divided into groups, because repeated divisions in different schools would increase events of chance. For that reason, a cloze test was used to match the subjects into two groups. Scores on the cloze test were ordered from highest to lowest in each class, and beginning from the top, blocks of two students were picked out, and then each randomly assigned to groups A and B. The units within each block would then be more homogeneous in their response than units selected completely at random. This procedure gives the name of the randomized block design, to be differentiated from the randomized group design (Edwards 1968, 156). An extreme form of the block design is the case in which each subject serves as his own control. In this design it is applied to both treatments being given to the same subjects. Figure 6 summarizes the design of the experiment, where M_t stands for matching, A & B for groups, W for written and O for oral modes, and t_1 and t_2 for trials.

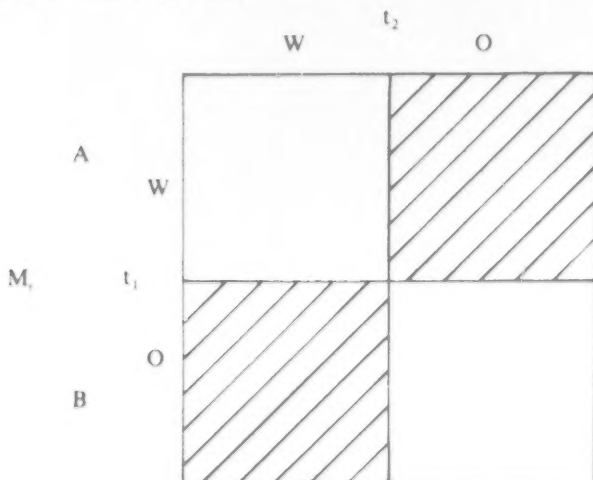


Figure 6. The design of the experiment.

This is called an orthogonal design. It enables collection of observations that will permit one to separately estimate and test for the various treatment effects as well as for interaction effects. The assumptions for this design are that (1) the observations within a given treatment combination are sampled at random and independently from a normal population, and (2) the number of observations in each possible combination of treatments is the same (Hays 1963, 393).

The second requirement will be met in administering the experiment. The first requirement will also be satisfied although the treatments in the first and second trials are the same, because, statistically, it is only needed that errors are independent. The issue will be discussed more thoroughly in the analysis of variance, further on.

The second main problem concerns the communicative test as a measuring vehicle. It consists of questions of validity, reliability, ability to differentiate students according to language proficiency, and administrability of the instrument in practice. These questions will be dealt with from different angles and through various means, but mainly by using statistical indexes and descriptions as aids.

The third main problem focuses on finding some tentative connexions between proficiency in a foreign language and some factors outlined in the instructional model of foreign language pedagogy. Connexions will be explored with the help of a questionnaire, given to the students participating in the experiment, and by handling the data through statistical means to find some regularities. Out of the factors represented in the instructional model, the object and discourse languages are measured through the communicative test. The questionnaire explores the instructional metalanguage, as indicated by the native/foreign language and by the instructional arrangements. The teacher's approach is indirectly measured by the students' answers. Affective factors among students, as part of the instructional model, will be rated on the basis of their motivation and their attitudes towards English studies.

8.3. Preparing the measuring instruments

8.3.1. The communicative test

Communicative proficiency in a foreign language has been defined as fluency, in different contexts (see Ch. 5). Two dimensions were distinguished in the foregoing conceptual analysis of fluency, which practically means that a person is capable of understanding messages and expressing understandable messages in a foreign language.

On the other hand, the person must master different contexts, which always have to be defined for the particular purposes. For this study, the context is defined in the curricular aims of the senior secondary schools

for the school-leaving phase. At the time of the study, schools followed the curriculum of *Nykykielet* (Modern Languages) (1971), which did not give a detailed description of the aims. It seems that the level of difficulty is primarily based on the experience obtained in examining students in the school-leaving examination, and consequently, the practical experience of this author was important in constructing the test. It is not to say that the matriculation examination does not measure whether the curricular aims have been achieved, but only that such aims have not so far been explicitly defined. This will hardly be possible in case of a higher-order proficiency in a foreign language. This is to defend the relevance of practice, in defining the linguistic aims of the curriculum.

In theory, the achievement of the curricular aims is evaluated in this study, but practically it is difficult to do it, because the measurement can only help to place students in an order of attainment, i.e. the achievement is scaled qualitatively, not quantitatively. In practice, the examiners always have to agree on the minimum level of passing the examination. As a consequence, the aims are basically taken into account in the choice of the text and its difficulty as well as in the art of measuring them.

The practical experience from a previous empirical study (Hellgren 1980) guided the preparation of the communicative language test. An interview offers natural colloquial material with sufficient subject-matter. So this type of authentic text was prepared for the test. Two similar parts of equal difficulty were needed for the test, and to standardize them in terms of the subject-matter, it was decided to interview one person and to restrict the scope of the topic to some extent and then cut the text in two.

The interviewee was an Englishman, a London secondary-school teacher, whom this author had personally met abroad. On the basis of this information, a framework of an interview was made out. The interviewer's questions were written out, and a few hints were given of what was expected as a response. The interviewee recorded the material in his home with his wife acting as the interviewer. The recording was later edited, in the sense that it was slightly shortened and made more concise with the outcome of two, fairly independent parts: the one mainly dealing with school life, and the other with university life in Oxford. Ten questions to be asked of students were prepared for each part, twenty in all. An attempt was made to obtain equally difficult questions for both parts.¹ The twenty questions about the text were read out on the tape by this author. The master tape for the communicative

¹ Editing the material and making the questions were done in cooperation with Mr. Juhani Kyllönen, M.A., an experienced English teacher at the senior level, and one of the markers of English essays for the matriculation examination. Mr. Eero Lehtonen, M.A., made the technical edition and improved the recording by reducing some background noise.

test was accomplished by copying the first part of the edited text straight through, and then recopying it by cutting the text into ten short passages, after which the relevant questions were inserted. A pause of two seconds was left between the text and the question to prepare the student for it. This had been considered important by the students in the previous study (Hellgren 1980). A pause of 40 seconds was reserved for the question to be answered. Then the interview went on till the next question, and so on. The second part was prepared in the same way. The total testing time was 31 minutes. The transcript of the whole text on tape is given in Appendix 1.

On the whole, the recorded version was reasonable, although it would have been better if the original interview had been made in a recording studio. On the other hand, in natural surroundings there is often noise in the communication channel, so that the hearer has to complete the missing parts of the message. This is made possible by redundancy of the language and with the help of the hearer's knowledge of the language. It may be unreasonable to require such a knowledge of a foreign language learner, but a slight noise should not prevent the message from going through.

The original idea was to pretest the material in a school but it was not possible to do quickly enough, and besides, a study on a small scale was not considered sufficient to allow for possible alterations. The previous study done by this author was considered adequate, particularly, since open-ended answers were used in the test.

8.3.2. The cloze test

For the experiment it was necessary to divide the classes into two equal groups. Since students vary considerably in their language performance, this ability was used as the basis of matching the two groups. A cloze test was chosen as a measuring instrument because it has proved a good and reliable technique of assessing integrated language proficiency in numerous studies (e.g. Oller 1973a). This author also had experimented with it on Finnish students (Hellgren 1976).

In the cloze technique, first developed by Taylor (1953), a test is prepared by mechanically deleting every n th word from the chosen extract. Subjects are required to fill in the gaps with a contextually appropriate word.

The face validity of the cloze test is good, as it can be said to correspond to a real communication situation with some noise in the channel to distort parts of the message. Thus, students and teachers may feel that the test requires relevant action by the language user. The content validity, in terms of the curriculum aims, has not been regarded as vital since the cloze test has frequently been used to measure global,

general foreign language proficiency. The view of language as a communication vehicle is essential to the construct validity. So is the internalization of a grammar of expectancy as an outcome of learning (Enkvist & Kohonen 1976). To be able to fill in such blanks, the reader makes use of the redundancy of the language. Psycholinguistically interpreted, the reader is able to anticipate the missing parts of the message with the help of a grammar of expectancy, which is an outcome of the adoption of the language.

Evidence of good empirical validity can be cited in the form of high correlations with other tests measuring language skills. The best correlations have been obtained between the cloze test and some tests of integrated language skills (Oller 1973a). A correlation of .82 has been found between the cloze test and dictation, and reliability was estimated at .80 (Oller & Conrad 1971). A similar correlation was obtained in another study, but the reliability index was .75 (Irvine et al. 1974). The cloze test correlated .79 with the standardized TOEFL Test Battery, by .76 with the listening comprehension part of the Battery (Irvine et al. 1974) and by .73 with the same LC test reported by Darnell (1968).

A correlation of .72 has been found between the cloze test and the school report marks of English among the students of the second form in a Finnish senior secondary school (Hellgren 1976, 48). On the whole, reliability indexes have been reported as varying .75—.80 (Oller 1973c), but the consistency of discrimination by items seems to be slightly inconsistent (Enkvist & Kohonen 1976, 22). The estimates of reliability obtained in Finland have been somewhat lower, at the level of .70 (Hellgren 1976; Kohonen 1976).

In marking the cloze answers two techniques of scoring have been used: (i) only the original word deleted has been accepted as a correct response, (ii) any contextually and semantically appropriate word has been accepted as a correct answer. The latter technique has frequently been recommended for foreign language testing (Oller 1973c; Irvine et al. 1974), because it is intuitively unreasonable to require the original word of foreign language learners. Furthermore, the contextual technique of scoring has made the variance increase and the average correlation of items with the total scores rise, resulting in a more reliable measuring instrument (Hellgren 1976, 44).

The extract for the cloze test was taken out of the structural exercises by Low (1979, 35—36). It was not used in its original exercise form with the gaps: instead, the text was completed and treated as a whole. From the text of about 250 words it was possible to make about 50 gaps if every fifth word were mechanically omitted. The lot for the starting point was drawn out of five, and it fell on the fourth word. From then onwards, every fifth word was deleted, and a total of 47 items was obtained (Appendix 2). A couple of times this mechanic rule was broken to avoid selecting too easy or too difficult a word. Although this was

against the theory of the cloze technique, preparing a good language test was considered more important than getting an instrument only to measure the redundancy of the text. Too easy or difficult items were to be anticipated on the basis of previous experience.

Since constructing a good cloze test seems to be somewhat of an empirical question, a try-out in a school was considered necessary. The test was given to all the students of the third form in Pirkkala senior secondary school ($N = 67$) at the beginning of September 1980. To obtain an index of empirical validity, the marks of English in the spring 1980 school reports were correlated with the scores, obtained in the cloze test. The results are displayed in Table 1.

Table 1. The results of the try-out of the cloze test ($N = 67$). (r_{tt} = reliability; r_{it} = mean item correlation)

Variables	Mean facility value (%)	$\bar{X}$	s	r_{tt}	r_{it}	Intercorrelation of variables
Cloze, 47 items	51.4	24.2	6.7	.82	.27	
Marks, 4—10 (4 = low, 10 = high)		7.0	1.2	—		.78

Since the test proved fairly reliable and capable of discriminating students, judged by the reliability estimate and the mean item correlation, there was no reason to change the test version.

8.3.3. The listening comprehension test

The original idea was to use a listening comprehension test as a matching vehicle as well. The listening comprehension test of English in the autumn 1980 matriculation examination was considered suitable for this purpose. Moreover, most senior secondary schools were likely to have the test tapes, provided by the Board of Matriculation examination for the students taking the test in the schools. Students in the third form of the senior secondary schools are usually given the autumn test as practice for the future examination. Consequently, it was practicable to collect the scores of the listening comprehension part.

The listening comprehension test consisted of 30 multiple-choice questions with four alternatives in each. The procedure was described above in Section 7.2.6.1. As the test is public after the examination and as it plays a minor role for this study, the text will not be reported here.

When the results of the listening comprehension test were studied, there seemed to exist some discrepancies in performance between it and the cloze test. It was thought that the listening comprehension test, as a multiple-choice test, probably gave a less reliable and more limited measure of a student's language proficiency than the cloze test. For that reason, the cloze test alone was preferred as a matching device. In addition to this, one school did not want to give their students the test until November, which further prevented it from being used for matching.

8.3.4. The questionnaire

The equality of oral and written performance was at stake in this experimental study. Consequently, it was necessary to control some background variables of students and teaching as well as the students' attitude to the test itself. A questionnaire was prepared for this purpose. It was originally made in Finnish, but a Swedish translation was prepared for the Swedish school. For this report, it was translated into English and is given together with the Finnish original in Appendix 3.

The questionnaire was actually planned to explain possible differences among oral and written performances but also, to account for proficiency as such. The questionnaire was planned along the following principles. Background variables were confined to concern sex, the amount of mathematics studied, and the number of years of English studies. These have been found of some importance to proficiency in a foreign language (questions 1—3). Motivation was measured by two components: firstly, as indicated by the student's willingness to devote his time to the use and studies of English (questions 4—9, 22). Out-of-class activities were considered relevant to this aspect. Secondly, questions 23—25 focused on motivation inspired by the English teacher and his behaviour. The student's attitude to English and to the aims of proficiency were rated by questions 15—20. The area was made to cover the pleasantness, difficulty, and importance of English studies.

The metalanguage of the instructional model was divided into questions 10—12 dealing with the technical arrangements of instruction, and into questions 13—14 dealing with the use of the native/foreign language in instruction. Lastly, the student's attitude to the communicative test experiment was assessed by questions 26—29. The scheme in Figure 7 gives an outline of the areas measured in the questionnaire, and the questions related to them.

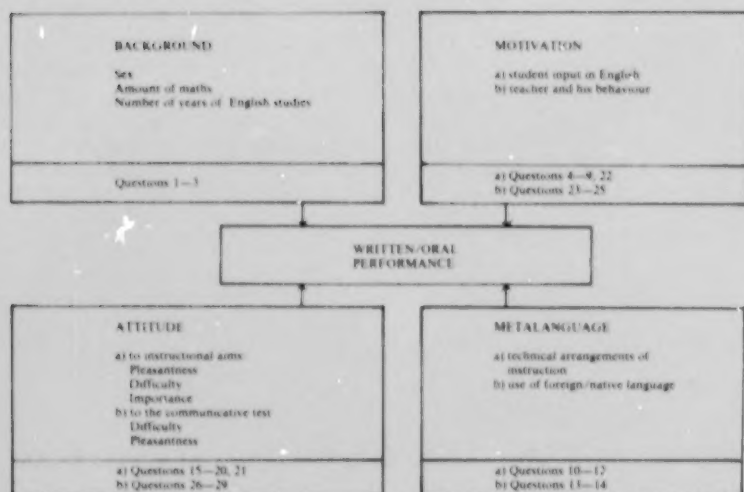


Figure 7. An outline of the areas measured in the questionnaire, and the questions related to them.

8.4. The population of the experimental investigation

The population of the empirical investigation is the students of the third form in senior secondary schools of Finland. In 1979, of all the students in the senior secondary schools 96.7 percent studied English as their first compulsory foreign language. The portion of girls was 61 percent, and five percent went to a Swedish school.

The empirical investigation was designed on the information of the records of the students in the second form, based on the situation in the autumn term of 1979, and reported by the Statistical Centre (Tilastotiedotus 1979). These students were to be in the third form the next year although their number would probably be smaller due to failures in the second form. Nevertheless, there was no reason to believe that the development should not be similar in all schools and the proportionate numbers thus remain the same.

Sampling students at random was impossible because of the character of the experiment and lack of resources. Besides, the characteristic class climate will be lost if students are not handled in terms of classes. The use of the class as a sampling unit is often an advantage, which is discussed e.g. by Alikoski (1969). An experiment in which several students' speech should be taped at the same time called for language

laboratories with tape recorders for students. These are called AAC laboratories to be distinguished from the AA laboratories, in which there are only earphones but no tape recorders in each booth. In Finland all schools do not have language laboratories. And those which have, do not all have tape recorders for every student booth in their laboratories. Consequently, the target population was reduced because individual testing would have been too cost demanding.

The frame population was confined to schools which had at least 15 tape recorders in their laboratories. There were 120 such schools in the whole country. The distribution of language laboratories in Finland is described in Table 2. The records are based on the situation in 1979. Thirty-one new language laboratories were to be installed in schools during 1980.

Table 2. Language laboratories in secondary schools by provinces¹ in 1979.

Province of	Language labs		Labs installed in 1980	AAC labs ²	Secondary schools	
	N	% of schools	N	N	junior	senior
Uusimaa	70	58	4	47	107	109
Turku and Pori	26	54	4	15	81	63
Häme	29	49	4	12	73	55
Kymi	21	54	2	9	41	31
Mikkeli	10	46	3	2	29	23
North Karelia	6	45	3	1	24	18
Kuopio	14	36	1	2	36	25
Central Finland	14	42	2	6	35	26
Vaasa	35	65	2	17	60	45
Oulu	30	52	4	5	59	41
Lapland	23	56	2	4	39	25
Total	278		31	120	584	461

¹ The province of Ahvenanmaa (Åland) is excluded.

² Labs with at least 15 tape recorders in the booths.

It is to be noted that many junior and senior secondary schools are placed in the same building and have their language laboratories in common use, which accounts for the discrepancy between the numbers and the percentages given in the above table.

The different provinces are approximately equal with more than 50 percent of their schools equipped with language laboratories. The 120 schools available for the experiment comprise 26 percent of the total 462 senior secondary schools in Finland. Strictly speaking, generalizations can only be made to the frame population, consisting of students in the 120 senior secondary schools with AAC laboratories. Nevertheless, there seems no reason not to generalize at least to students in the senior

secondary schools equipped with language laboratories. This is because the most decisive factor for language learning is whether the teacher uses the laboratory efficiently or not. The quality of the equipment in this sense is less important. Accordingly, the results of the investigation can be based on about half of the target population.

8.4.1. Sampling technique

In this study, the primary interest is in a student whose performance is measured and who thus stands for an elementary unit. Since simple random sampling was impracticable and did not serve the purpose, a senior secondary school was selected as a sampling unit. The 120 schools with AAC laboratories were unevenly distributed in the whole country, so that a simple random sampling out of them would have favoured certain areas. The interest in students, not in the language laboratories of different areas, compelled the use of a stratified sampling to guarantee a proportionate sample of the entire frame population for control of plausible regional differences.

The country was divided into five sectors according to the provinces. The sectors differed in numbers of students and senior secondary schools, which, consequently, were used to obtain probabilities proportionate to size. Ten senior secondary schools were expected to be within resources. Thus the rounded probabilities proportionate to the total number of schools give the number of schools to be randomly drawn for the sample in each sector. Rounding makes this a little less precise technique, but it is hoped the probabilities of students will level it.

In practice, intact classes must be handled, which prevents the investigator from using the exact probability numbers. Long and short courses in mathematics may cause a systematic difference between classes. That is why it was necessary to sample at least two classes in each school. The final number of students in the sample was designated to be 400. It was considered to be the lowest limit, because the sample had to be divided into two groups. Although the figure is arbitrary, it was the result of the following reasoning. Since the students taking the long and short mathematics courses are usually in separate classes, two classes had to be chosen in each school. An average number of students is 30 per class. Since the research was done in phases, this would bring about a likely dropout of students, due to absence from school. The figure was estimated at ten students per class, which meant that at least forty students had to be obtained from each school. The multiplied number then resulted in 400. Of all the students in the second form 38.6 percent studied a long course in mathematics versus 61.4 percent with a short course. An attempt was made to preserve these proportions in the sample to be drawn. Table 3 summarizes the stratifying and sampling design.

Table 3. The sampling design. Based on the records of the second-form students in the senior secondary schools from the autumn term of 1979.

Sectors/Provinces of	Population						Sample		Population		
	Number of					Percentage of		Number of		The average number of students per	
	students	schools		classes		schools	students	schools	students	school	class
1. North Finland	5552		66		192						
Oulu	3566		41		125						
Lapland	1986		25		67						
2. East Finland	7678		97		265						
Kuopio	2078		25		70						
North Carelia	1440		18		49						
Kymi	2475		31		86						
Mikkeli	1685		23		60						
3. Central Finland	5323		71		187						
Central Finland	2005		26		68						
Vaasa	3318		45		119						
4. West Finland	9851		119		337						
Häme	4700		55		161						
Turku and Pori	5078		63		173						
Ahvenanmaa (Åland)	73		1		3						
5. South Finland	8268		109		280						
Uusimaa	8268		109		280						
Total	36672		462		1261						

In the above table the average number of students per school varies somewhat from sector to sector. Instead, the average per class is approximately the same for all the five sectors. As this is a more important factor for oral language learning, the schools might be regarded as equal in this respect.

8.4.2. Sampling of schools

The names of the 120 senior secondary schools were each written on a separate card. The cards were divided into groups of the five sectors of the country. The cards in each group were thoroughly mixed and schools were randomly drawn out of them according to the sampling design. For each group a double set was picked out and recorded in the drawn order for possible later use to compensate a loss. Table 4 records the sample of schools. A map of Finland in Appendix 4 displays where the schools are located.

Table 4. The drawn sample of senior secondary schools.

Schools	Province of
I North Finland	
1. Kuusamon lukio, Kuusamo	Oulu
II East Finland	
2. Kouvolan lyseon lukio, Kouvola	Kymi
3. Ruotsinsalmen lukio, Kotka	Kymi
III Central Finland	
4. Pietarsaaren lukio, Pietarsaari	Vaasa
5. Kirkkopuistikon lukio, Vaasa	Vaasa
IV West Finland	
6. Svenska samskolan (Swedish), Tampere	Häme
7. Lammin lukio, Lammi	Häme
8. Paimion lukio, Paimio	Turku and Pori
V South Finland	
9. Korson lukio, Vantaa	Uusimaa
10. Vehkojan lukio, Hyvinkää	Uusimaa

To get an idea of the representativeness of the sample Table 5 describes the distribution of schools by the type of communities and the corresponding figures in the drawn sample.

Urban schools seem to be more frequently equipped with AAC language laboratories than rural ones. The figures in the drawn sample are proportionate to the sampling population.

Table 5. The distribution of senior secondary schools by the type of communities.

Sector/ Province of	Number of schools by communities			Schools with AAC labs			Drawn sample		
	Urban	Other	In All	Urban	Other	In All	Urban	Other	In All
I North Finland	27	39	66	7	2	9		1	1
Oulu	15	22		3	2				
Lapland	4	17		4	—			1	
II East Finland	52	45	97	12	2	14	2		2
Kuopio	11	14		1	1				
North Carelia	8	10		1	—				
Kymi	22	9		9	—		2		
Mikkeli	11	12		1	1				
III Central Finland	27	44	71	13	10	23	2		2
Central Finland	9	17		4	2				
Vaasa	18	27		9	8		2		
IV West Finland	71	47	118	20	7	27	2	1	3
Häme	35	20		10	2		1	1	
Turku and Pori	36	27		10	5		1		
V South Finland	94	15	109	43	4	47	2		2
Uusimaa	94	15		43	4		2		
Total	271	190	461	95	25	120	8	2	10
%	59	41	100	79	21	100	80	20	100

8.4.3. Sampling of students

To control equality of oral and written answers to the same questions, it was necessary to match the students. A so-called randomized block design was used in matching. In forming blocks it is anticipated that, in the absence of treatment effects, variation within each block will generally be less than would be found if the same number of observations were obtained from subjects selected completely at random. If there is substantial variation between the blocks, the analysis of variance effectively eliminates this source of variation from the estimate of experimental error. (Edwards 1968, 171).

Since there was good reason to believe that students varied considerably in their verbal ability, this variance was controlled by forming blocks on the results of pretesting. A cloze test and a listening comprehension test were used. As was previously described, the cloze scores of each class were arranged in order from highest to lowest. When several students obtained the same score, they were ordered with the help of the listening comprehension test. Blocks of two students were then systematically picked out, starting from the highest score. By tossing a coin, the students were then randomly inserted into groups A and B, which thus formed a block of a matched pair. The procedure was repeated for each class because intact classes had to be handled. The actual division of students into the two groups took place in the experimental process, and thus the students and their teachers did not know about their placement prior to the experiment.

8.5. Data gathering

The cloze test was sent to the schools at the beginning of September 1980. The teachers were allowed to give the test at an appropriate time. A 45-minute lesson was reserved for that, although 30 minutes were considered enough. As the official matriculation examination of listening comprehension for English was held on the 11th of September, the teachers were asked to give the test to their pupils at a convenient time within September. The results of this listening comprehension test, together with the cloze test papers, were returned to the researcher to be marked. This was done as planned except that School 9 did not take the listening comprehension test until November, which also prevented the test from being used for matching as planned in advance.

The teachers did not mark the cloze papers but instead, were given a key to the right solutions to be discussed in class. In December 1980, they were informed of the results of the scores with a suggested mark, obtained on the basis of the calculations from the whole data, with the pretesting included. The cloze papers were left with the researcher for

further analysis.

The actual experiment was carried out in the different schools during the period of October 30 to December 5, 1980. To equalize the test situations, it was necessary that the experiment be made by the researcher himself, who would then know what actually happened and how to react to unexpected events because only he knew what was crucial for the research. A convenient date was agreed on with the teacher on the telephone, and the times ran smoothly according to the timetable.

Before the actual testing, the researcher's cassettes or open reel tapes were inserted in the tape-recorders of the pupils' booths. When the students entered the classroom, they were divided into two groups on the basis of the cloze test results. The answers to questions 1 — 10 of Group B were recorded while Group A wrote their answers to the same questions. After this the groups changed places and the oral answers of Group A to questions 11—20 were recorded, whereas Group B wrote their answers to the corresponding questions on sheets of paper.

At the beginning of the session the instructions were read out in English and explained in a few words in Finnish. The procedure of the listening comprehension test in the matriculation examination was quite familiar to the students, and this test was run accordingly, with the exception of open answers and the questions heard on tape. Consequently, hardly any disturbance was caused by not knowing what to do during the test.

The first section of the text was listened to straight through, and at the beginning of the second listening, the experimenter put on the pupils' tape-recorders through the teacher's control desk. The tape-recorders were not stopped until the end of the whole part. From previous experience, it was considered too risky to stop the recorders after each answer, because some cassettes or tapes might not start again if they were stopped at intervals. Thus, the text was also recorded on the pupils' tapes. This was possible because the total use of time allowed two recordings on the same track.

This experiment completed, the students were asked to fill in or choose the right alternatives in the questionnaire distributed to them. In the meantime, the tapes were rewound, collected and new ones inserted in the tape-recorders or cassette-players for another session. This is how the procedure was repeated in every school.

As a rule, the classes followed one after the other with a few exceptions: In School 1, one group came two hours later than the two others; in School 3, the second group came one hour later, and in School 8, there was an hour's interval between the sessions. This may hardly have any considerable effect on the results. Since the questions demanded one's own production, the answers were not easily reported.

There were some disturbances during the testing. During the second

session in School 1 some cassettes stopped in the process of recording. Attempts were made to get them restarted. A couple of them did not, however, work. In School 8 the problem reoccurred. The laboratory equipment caused another problem after the testing. A number of cassettes were found unrecorded, which reduced the number of subjects. Since the misfortune was casual, the effect is likely to be slight on the outcome of the experiment, particularly, as the schools were not employed as the object of comparison. Due to absence from school, missing information or matches, and some unrecorded tapes or cassettes, the remaining number of the subjects was 406.

8.6. Scoring the instruments

8.6.1. Scoring the communicative test

Evaluating the open answers of the communicative test can be seen as an analogy to doing hermeneutic research, i.e., the scorer attempts to understand the pupil's intentions and aims. Consequently, he is apt to understand the message correctly despite some linguistic or grammatical errors. This is especially facilitated by the background information of the message, i.e. the context of the discourse.

A mechanic scoring procedure is claimed to add to the objectivity of the test. This is only an illusion since hardly any test on human abilities is free from subjectivity. It only manifests itself in different phases — either in constructing or in scoring a test. In this experiment, subjectivity was centred upon the scoring. Consequently, it was necessary to find as clear principles as possible to control fluctuation in scores. Such principles must be seen from the communicative point of view. They are partly derived from the theoretical framework presented above, and partly from the experience obtained in marking the answers in a previous study (Hellgren 1980). The total of four points was awarded for an answer to one question, distributed according to the principles explicated below.

As communicative foreign language proficiency can be said to consist of two aspects, the communicative language use and the mastery of the integrated language, these sides have to be considered when appraising the responses of the test. The answers must be evaluated within the framework of the test context. This may sound self-evident, but it is not so in case of errors and mistakes. If the answers were viewed as isolated sentences, a semantically wrong word or a word in a slightly distorted form might cause a complete breakdown in communication. Nevertheless, when the scorer is aware of the background information — the text and the question — he can conclude, i.e. understand the answer better, in spite of its slight distortion. An example: *It is polished pro it is abolished.*

Following this principle can be justified by referring to a normal communication event. The hearer has a wider context available than a single utterance. On the other hand, this communicative aspect adds to subjectivity because understanding is dependent on the hearer/reader, his knowledge and state of mind. There is hardly any objective way of deciding what utterance mediates the message correctly. A kind of "normal" interpretation might empirically be found through an average measure from many different scorers, but it offers no guarantee of objectivity. On the other hand, the test used in this experiment reduces subjectivity even to a greater extent than in a real situation of communication, since the background information is given, and thus largely shared by the communicating partners.

The first of the two aspects, communicativeness, is considered when marking the answers, the correctness of the response is decided upon, in terms of the contents. If the answer is correct and complete enough, a minimum of half of the total score, i.e. two points, was given provided that language errors or mistakes did not prevent the scorer from understanding what had been meant. If there was a blank, or a word in the mother tongue inserted, this was interpreted as a breakdown in communication.

Another factor of subjectivity comes from the interpretation of what and how much of the contents is required for a complete answer. This can be decided upon question by question, but it shows that items are not quantitatively alike, i.e. the language proficiency cannot be measured in absolute quantitative terms, but the quantities are relative. The relative amount of the information determines the scores given for a response. A fairly reliable measure may be obtained by comparing the responses to a single question, but the criteria vary from question to question.

Evaluating only the communicative aspect of the responses would obviously lead to a rather rough categorizing of answers, if their linguistic quality is not taken enough into account. Therefore, it was necessary also to appraise the utterances from that point of view. This meant scoring according to the correct usage of the language.

Spelling mistakes as well as errors in pronunciation, which did not cause a change in the sense of the word, were omitted. Also, due to the character of the test, missing articles or the addition of an extra one was ignored, except in the case of very short and weak answers. This principle can be defended by the quantity: If one pupil uses more language than another, he is likely to make more mistakes, at least slight ones. A weak pupil very often gives short answers, either because of his incapability of expressing himself in the foreign language, or for fear of making mistakes. Anyway, his communicative proficiency in the foreign language is limited. Another justification for our procedure is that in oral answers, the pupil may utter extra sounds which fill in the time gap

of thought or hesitation. Such sounds may or may not be articles. Because of the difficulty of interpretation, they were not observed.

For each error and mistake in grammar, morphology, semantics and syntax, one point was reduced, but not more than two points in all. This range may seem too narrow, but it has been shown to work very well in the previous study, and also in this one. It is accounted for by the fact that semantic mistakes often cause a hindrance in understanding, which means that such mistakes are interpreted as breakdowns in communication. As it is not easy to separate these aspects completely, a total score for an answer allows for judging a response holistically, and then reducing the necessary minus points. This procedure is considered to give a more reliable score than keeping the two aspects completely separate from each other.

Also, the situation becomes more complicated in case of a long and complete answer with an error of morphology or syntax. In this experiment, the rule was followed that a long and fluent, original answer was credited with four points although a slight grammar error might have occurred there. The communicative aspect also leads to a "softer" view of the answers, in the sense that a given response, seriously wrong as it may be, must be credited with a score if it has an element of truth and correctness. This may be justified by the mutuality of the communicative situation. The speaker and hearer can always correct each other's utterances or make them more precise.

The answers were marked by the researcher. The written responses were scored first. All the answers to the same question were scored at the same time to facilitate a comparison among them and to guarantee their equal treatment. Certain answer types were written down under the expected answer to the question, and the score given was put down beside the response. Thus the scorer could see at a glance how different types of answers with various errors were scored. No markings were made on the pupils' answer sheets but their scores were directly coded in the raw score matrix. The written answers were first marked to obtain a marking framework for the oral answers, which could not be scored by listening to all the individual answers to the same question. All the responses of one pupil had to be listened to successively. It would have been quite impracticable to change tapes after every answer. The estimated time spent on a tape, with two students' answers on it, was about 30 minutes. This, adding two minutes for rewinding and handling the tape, would come up to about 127-130 hours. The actual time used was higher because it was necessary to have breaks between the listening periods. Also, some parts had to be replayed to hear what was said.

The oral responses were scored by listening to them, but they were also transcribed to facilitate the control. After the first scoring of all the answers — written as well as oral ones — the marking principles were consulted and adjusted, all the responses underwent a thorough check to

enable an equal marking throughout. A more detailed list of observations on the answers per question is given in Appendix 5.

The original idea was to let another scorer mark the answers, but later on, the conclusion was drawn that the procedure would not be any more reliable than, if two randomly chosen scorers were marking the same answers. In other words, such an empirical way of measuring reliability would be too haphazard to be worth a try. For this experiment, the vital reliability was to control the fluctuation of scores to the same question from answer to answer. The procedure of writing down many model answers, already scored, as well as the transcription of the oral responses, together with the after-check of the scoring, were considered sufficient to guarantee a controllable reliability. The subjectivity of marking the answers would thus spread over all scores, and equally influence both the written and oral responses.

8.6.2. Scoring the cloze test

The cloze test was scored contextually, just as in the pretest, i.e., all syntactically and semantically suitable alternatives were accepted as correct, and awarded with one point. In case a word was missing, or an incorrect response offered, or more than one word used, no point was given for the item. A marking list was kept on the alternatives given by different students, and the scores given for them, to control equal points for similar responses. Here, too, attention was paid to the consistency of scoring. Just what was said about using another scorer previously, applies to the cloze test scoring. It is a very contingent matter to have just one other scorer. It deserves a study of its own to find out what points different scorers disagree on. This was not possible, and not worthwhile, for this study.

8.6.3. Scoring the questionnaire

The questionnaire was scored as follows (see Appendix 3). In questions 1—2 the alternative *a* was coded for 1, and *b* for 2. In question 3 the number of years was directly coded. There was some hesitation among students about including the current year or not. To equalize interpretation, they were asked to include the current year. The question would have been clear at the beginning of the school year, when the test was originally to be performed.

In question 4 the negative answer gave one point and a positive answer the following scores: 1 p. no—2 weeks; 2 p. 3 weeks to 1 month; 3 p. 2 months; 4 p. 3 months; 5 p. more than 3 months. There are not usually long courses organized for school pupils, but the idea was to count all the different courses together.

Question 5 was scored: 1 p. No—2 weeks; 2 p. 3—5 weeks; 3 p. 6—9 weeks; 4 p. 10—13 weeks; 5 p. more than 14 weeks. The aim was to distinguish long-term stays from short ones. The highest score 5 was normally used for one year's stay in a foreign country.

In questions 6—8 the alternatives were scored *a*: 1 p., *b*: 2 p., *c*: 3 p., *d*: 4 p., *e*: 5 p. The alternative *f* of the eighth question had not been used. In the ninth item, an open-ended answer was offered, to map various alternatives that students might have made use of. The question concerned the improvement of the English language through out-of-class activities. There were mentions of listening to music, and its lyrics, penfriends, films, TV programmes, exchange students, summer courses, language cassettes, language clubs, foreigners, friends. As it was not considered possible to estimate their value in scores, coding was done as follows: 0 if there was a gap, 1, if one thing was mentioned, and 2, if two things were mentioned. So the range was 0—2. This principle was regarded as relevant because, basically, the question concerned the subjective view of the improvement of one's language skills, and it was brought forth through this scoring.

The rest of the questions were provided with 5 alternatives. Since the positive and negative ends varied in the question forms, the dimensions were, however, coded similarly so that the highest score of 5 was given to the most positive attitude, with easiness included. Consequently, the coding was as follows:

In questions 10—11, 16, 17, 19, 22, 24:

a 1 p. *b* 2 p. *c* 3 p. *d* 4 p. *e* 5 p.

In questions 12—15, 18, 23, 25—29:

a 5 p. *b* 4 p. *c* 3 p. *d* 2 p. *e* 1 p.

8.7. Results

8.7.1. The first problem area

To facilitate the following discussion, the experimental groups are talked of as A and B. Group A took the first part of the test (questions 1—10) by writing their answers, and the second part (questions 11—20) by giving their responses orally, whereas Group B took the first part orally, and the second part by writing their answers. The main results of the experimental tests are displayed in Table 6, and a detailed item analysis is given in Appendix 7.

Table 6. The main results of the communicative test as four separate subtests, of the cloze test and of the listening comprehension test. $N = 406$, $n_A = 203$, $n_B = 203$.

Tests	Maximum score	%	$\bar{X}$	s	r_{11}	r_{12}
Communicative test						
Subtests						
Written response: items 1—10,						
Group A	40	57.6	23.0	8.7	.81	.49
Oral response: items 1—10,						
Group B	40	57.8	23.1	9.2	.83	.53
Written response: items 11—20,						
Group A	40	43.1	17.2	7.2	.78	.46
Oral response: items 11—20,						
Group B	40	42.4	17.0	5.0	.83	.53
Cloze test, Group A	47	54.0	25.4	7.1	.85 ¹	.30 ¹
Group B	47	53.6	25.2	6.8	.85 ¹	.30 ¹
Listening comprehension test						
Group A	30	59.7	18.0	5.1	.78 ²	.38 ²
Group B	30	60.0	17.9	4.8	.78 ²	.38 ²

¹ Estimates computed over all subjects ($N = 498$)

² Estimates computed for the matriculation examination in autumn 1980 ($N = 1740$)

The average facility value was obtained by the percentage of correct answers averaged per test. The values as well as the means seem to be similar across the response modes, but not across both parts of the communicative test, the former part being somewhat easier than the latter. Both groups performed equally well in the cloze and listening comprehension tests. The variances in the listening comprehension test are lower than those of the other tests. This is probably due to the multiple-choice type of the test. The reliability estimates of the listening comprehension test could not be calculated because of the lack of scores per items, so that the outcome obtained in the matriculation examination is reported in the table.

To test the hypothesis of the first main problem, an analysis of variance was needed. The effects of oral and written performances can be considered a fixed treatment (t_1 in Figure 6, page 95), because inferences are to be made only about differences between the two response modes and not about other treatments that might have been included (Hays 1963, 357). The other treatment factor (t_2 in Fig. 6) can be described as a content treatment, in the sense that contents of the two parts of the test may differ in difficulty. The two parts can be viewed as only two factor levels, forming a random sample out of the whole range of possible content levels. Thus, many more levels of the factor are possible than actually occur as observations in the experiment. So the content treatment (t_2) is a random factor in the experiment. Such a

factor combined with a fixed factor calls for a mixed model of analysis of variance, which is a combination of the fixed effects and random effects models. In the mixed model, each individual observation results in a score that is a sum of both fixed and random effects (Hays 1963, 439).

If under the fixed effects model inferences made concern means and differences between means, inferences made using the random effects model, deal with variances, i.e. the basic inference has to do with the variance of the population of effects actually sampled by the experimenter (Hays 1963, 427—428). Although the computational procedures are the same under different models, it makes a difference in the *F* ratios employed to test for the different effects. Also, assumptions behind the analyses differ.

In spite of the model of analysis found, there are two important points to consider. Firstly, there are two treatments given to the same subjects, and secondly, the groups, A and B, are matched. No direct design for analysis of variance has been found in statistical literature (Lindquist 1953, Hays 1963, Edwards 1968, Winer 1971). The works consulted discuss an analysis of variance for repeated measures or matched groups, but not by including both procedures in the same design.

The crucial point was thought to be that the same pupils were observed twice. The advantage gained in an experiment having repeated measures lies in the fact that the pooled within-person variation can be divided into two nonoverlapping parts: one is a function of experimental error plus the main effects of subjects within groups (i.e. individual differences). The other part is a function of experimental error and $t_i \times$ subject-within-group interaction. (Winer 1971, 520). If there is some interaction, it can then be reduced from the second part of the within-cell variation, and thus it is possible to lessen error.

Worth noting is that the fixed effects factor is tested against interaction in the mixed model (Hays 1963, 444). There is no complete agreement about this procedure, but because of the repeated measures in this experiment, this more robust error term was used for caution. On the other hand, the hypothesis associated with the random effects factor is tested against means square error in the *F* ratio. The test for interaction in a two-way analysis is the same for all three models: $F = MS \text{ interaction} / MS \text{ error}$. (Hays 1963, 444).

Verbal ability was regarded as a nuisance variable, and therefore it was introduced into the experiment as an experimental factor. Thus, each level of verbal ability was represented by the matched pairs of the subjects. Strictly speaking, all possible levels of verbal ability are not represented but rather, a random sample of them. This randomness was preserved by using the randomized block design in matching. The matching procedure gave the advantage that variation between groups A and B was eliminated, and thus the variance considered as error would be smaller, which again added to the precision of inferences. The

dependence brought about between scores will not, however, cause any dependence between the error portions of the scores between the groups. Such independence is the basic assumption behind an analysis of variance (Hays 1963, 456). Errors in the scores of repeated measurements on the same subjects must also be independent.

Matching students was a way to control carry-over effects. It was supposed that possible effects like that would similarly affect the oral and written performances of the latter part of the test. Matching enabled the study of the interaction effect, and thus, estimate independence of errors.

Recently, time series methodology has been attempted and employed for situation that preclude the use of standard linear model designs and that violate conventional statistical assumptions. One example of such methods is a restricted factor analysis (Jöreskog 1979, Störbom 1979), using e.g. so-called maximum likelihood estimations in analyzing the data. Unlike the classical factor analysis model, there is no assumption that all error variates are uncorrelated. Since such designs are still vulnerable to internal and external validity problems and since computer programmes for them were not easily available, it was not thought worth the effort to try to analyze our data with such methods.

The last thing to note about our analysis is its degrees of freedom. For a randomized group design, the number of degrees of freedom is $2(n-1)$ for the test of significance, whereas only $n-1$ d.f. are necessary for the randomized block design (Edwards 1968, 166). In our case there is a difference of only one degree of freedom. The variance analysis for the experiment is displayed in Table 7, page 118.

The F value for the difference between the response modes is .125 and nonsignificant, so that the null hypothesis of no difference cannot be rejected. This means that a difference between oral and written response modes is very unlikely to occur among the population, especially, since a relatively large number of subjects is employed in the test of significance. On the other hand, the two parts of the communicative test seem to differ from each other, since the F value for the difference is statistically highly significant. This may be accounted for by the contents, causing more difficulty in the second part. It is not of great importance to the experiment. A more vital thing is whether the interaction effect is significant. The F value of .256 for the interaction of trials and response modes is nonsignificant. Consequently, we have a broader basis to generalize, that using oral or written foreign language in answers shows equality in level of performance in spite of variation in difficulty. This certainty is due to the fact that the main effect of treatment, has been tested in conjunction with two variation levels of the contents factor without having held it constant at an arbitrary level. For the hypothesis testing of a unitary factor of foreign language, the result means that the experiment did not succeed in finding proof for rejecting the theory of a unitary factor.

Table 7. Analysis of variance for the experiment over groups and treatments, and by groups.

Source of variation	Sums of squares	d.f.	Mean squares	F
Over groups and treatments				
Between subjects	48 653.1	405		
response modes	.56	1		.125
error	48 652.5	405	120.13	
Within subjects	14 325.1	406		
trials	7 186.75	1	7 186.75	408.11**
trials $\times$ response modes	4.5	1	4.5	.256
error	7 133.81	405	17.61	
Total	62 978.1	811		
Over Group A				
Between subjects	22 875.2	202	113.24	
trials	3 414.44	1	3 414.44	238.9***
Error	2 885.87	202	14.29	
Total	29 175.5	405	72.04	
Over Group B				
Between subjects	25 777.3	202	127.61	
trials	3 776.81	1	3 776.81	179.59***
Error	4 248	202	21.03	
Total	33 802.1	405	83.46	

Inferences made above were made on the basis of variances. It is, however, important to know whether the trends are also equal over the response modes, which can be studied by correlation coefficients. Since the communicative test was divided into two subtests for the two groups, and since they were matched into A and B, it was possible to calculate intercorrelations for the subtests and groups. They are reported in Table 8.

Table 8. Correlations between tests. (Decimal points omitted)

Variables	2	3	4	5	6	7	8
Communicative test							
1. Written, items 1—10, Group A	72	52	54	69	69	64	43
2. Oral, items 11—20, Group A		51	58	67	66	67	54
3. Oral, items 1—10, Group B			75	63	65	50	66
4. Written, items 11—20, Group B				67	70	53	70
5. Cloze test, Group A					94	64	66
6. Cloze test, Group B						62	66
7. Listening comprehension, Group A							51
8. Listening comprehension, Group B							

The correlation coefficients between the written and oral response modes for groups A and B are .72 and .75, respectively, which shows a strong linear connexion between the response modes. The squares of the correlations, the coefficients of determination, are .52 and .56 for the groups A and B, respectively, and inform us of the "goodness" of the linear rule for prediction, i.e. how well we can predict the scores of the oral or written performances when one of them is known. A small negative effect on the correlation indexes may be due to the fact that scores among the first and second parts of the test distributed slightly differently (see Appendix 6). A fairly strong linearity is also seen between the same tasks answered in the different modes when calculations are made over the matched subjects: .52 and .58 for parts I and II, respectively.

The reason why the above correlation indexes were not even higher may be that some error was caused by the immediate repetition of the measurements, the small number of items, or the restriction of range in the second part of the test, due to its difficulty. Also, the subject-matter may have brought about differences in performance which were not linearly dependent on language proficiency.

8.7.2. The second problem area

The second problem area deals with reliability and validity estimates as well as the ability of the test to differentiate subjects according to proficiency in a foreign language, in this case English. There exist many theories about measurement. When achievement is measured, they are called test theories. If the scores obtained are assumed to represent true scores within the limits of error, a true score theory can be applied, whereas a latent trait theory explains the scores which are supposed to be the result of a latent trait. The approach of the true score theory is person-oriented: a person's performance is supposed to directly reflect the trait measured. The approach of a latent trait theory is reaction-oriented since the scores are assumed to be a result of the interaction between a certain latent (nonobservable) trait and an item. (Konttinen 1980, 22—23). Although the latent trait theory has been developed since the 1940s, its everyday applications have not been many until recently. Nevertheless, the following analyses are based on the classical test theory. It was considered justified because different tests were used, in which interest was focused on the subjects' performance, rather than on the interaction of an item of a nonobservable latent trait.

The average facility value can be regarded as an index of difficulty level. Although appropriate limits cannot be given for a statistically good test, an intuitive impression of an achievement test is that it should not fail to come up to at least a minimum of 50 percent correct

responses. The second part of the communicative test did not quite meet this demand. As previously reported, the counterparts of the subtests did not differ in difficulty, but the successive parts did. Difficulty or facility does not necessarily influence the test qualities.

A more important thing for a test is that it is reliable. Reliability concerns the extent to which measurements are repeatable — by the same individual using different measures of the same attribute or by different persons using the same measure of an attribute (Nunnally 1967, 172). In the classic test theory, measurement error is seen to be composed of a random sample of items from a hypothetical domain of items. Coefficient alpha developed by Cronbach (1951) is often used to estimate reliability in such cases. It can be regarded as an internal consistency estimate of composite test reliability. The reliability indexes reported in Table 6 are good, in spite of the low number of items for each subtest. Increase in reliability was studied by adding items: the oral and written subtests were combined for the groups separately to obtain the indexes of reliability shown in Table 9.

Table 9. Increase in reliability by combining the response modes for the groups, and by combining the different response modes over the groups, separately.

Groups/Modes	Items		
	1—10	11—20	1—20
A	.81	.78	.89
B	.83	.83	.89
Written			.86
Oral			.88

By making the test longer, reliability seems to improve, but perhaps not so markedly as one might expect. An essay-type scoring gives a fairly reliable measuring instrument, even with few items. Reliability was also assessed by calculating estimates for the answer modes, when students from both groups influenced the instrument. Some improvement was to be seen, but it was not considerable.

The tests performed by the two groups can be interpreted as parallel tests. The approximately equal means and standard deviations by items for the two groups (Appendix 7) clearly demonstrate that the same measure of language proficiency in another group produces a repeated result. The similar indexes of coefficient alpha also indicate this.

Test homogeneity has been discussed at length in the test theory literature. There is no general agreement as to just what this term should mean and how homogeneity should be measured (Lord & Novick 1968, 95). Coefficient alpha has often been cited as an index of test

homogeneity. Green et al. (1977) maintain that the concept of internal consistency and the concept of homogeneity should not be mixed. They show that it is possible to obtain high internal consistency as measured by coefficient alpha when there is no general factor running through the test items. The chief defect of coefficient alpha as an index of homogeneity is its susceptibility to increase with increases in the number of items.

Homogeneity of the communicative language test was investigated by studying the interitem correlations, and then factoring the subtests for the groups separately. The interitem correlations are reported in Appendix 8. On the whole, all items correlate neatly with each other, not very strongly, but fairly well. Only a few low correlations (below .10) occur, and the most frequent of them appear among items 13 and 16. This may be due to the difficulty of the items. A restriction of range and some chance may have caused the low correlations. Also, some aspect other than language proficiency might have been involved in the measurement.

Inasmuch as we were concerned with the analysis of the total communality rather than the total variance of the items, the principal axes model was used for factoring, i.e. the numbers placed in the diagonal of the correlation matrix were communalities, and not unities, which would have led to principal components. The factorization was performed in two phases. First, a structure of four factors was attempted. Since the first factor gathered most of the communality and the other factors only a few percent, a new attempt was made for a solution of three factors.

Researchers apply various methods and different criteria for the decision on which factor loadings are worth considering when it comes to interpreting the factors. We might have contented ourselves with only one factor, because the eigenvalues were below a unity after the first factor had been extracted (Kaiser's criterion). Another criterion is that an acceptable level of significance may be applied just as to correlation coefficients. A disadvantage is the absence of any adjustment for the number of variables, or the factor under consideration. It has been shown that as one progresses from the first factor to higher factors, the acceptable value for a loading to be judged significant should increase. The gradual intrusion of unique variance into later factors requires some adjustment in the level of significance. The Burt-Banks formula allows not only for the sample size, but also for the number of tests correlated and the number of factors extracted up to and including the one under examination. (Child 1970, 48).

The Burt-Banks formula was used as a criterion. The formula and its application to this data are given in Appendix 9. Since the other two factors contained statistically significant loadings after the first factor had been extracted, they were taken along to enable a broader

comparison between the factor patterns and structures of the subtests. If there arose some disagreement between them, it might be explained by the different response modes. Table 10 reports the eigenvalues and the percentages of the variances explained by the factors. The detailed solutions are given in Appendix 9.

Table 10. The principal axes solution of factors for items as subtests of the response modes.

Factors	Items 1—10			Items 11—20		
	F I	F II	F III	F I	F II	F III
Eigenvalues						
Written	3.18	.47	.26	2.99	.31	.27
Oral	3.51	.29	.28	3.69	.38	.27
Percentage variance						
Written	31.8	4.7	2.6	29.9	3.1	2.7
Oral	35.1	2.9	2.8	36.9	3.8	2.7
Cumulative percentage variance						
Written	31.8	36.5	39.0	29.9	32.9	35.6
Oral	35.1	38.1	40.9	36.9	40.7	43.3

Since the first factor in each solution extracts most of the variance, the measuring instrument seems to be homogeneous to some extent. However, only about 30—32 percent of the total variance is explained in the written subtests, whereas in the oral subtests the first factors gather 35—37 percent of the variance of the items. The slight but consistent differences between the written and oral subtests may be due to the response mode although an explanation to it is difficult to give. An interesting feature is that the first factors of written performance share less common variance of the variables than those of spoken performance. The total variance of the common factors does not show the same difference between the response modes in the first part as in the second part of the test. There may be some features specific to written mode which do not emerge in oral performance.

The items of the communicative test did not produce one single factor, but two additional minor factors. This was to be expected because the questions were formulated so that they would cover a varying amount of information, or require partly repetitive reaction or inference-making. Thus, different human processes and factors would be involved. The consequential shortcoming is that it is not possible to interpret the factors explicitly. For that reason, further investigation is here called for.

Error has been somewhat reduced in the second part of the test, and this can be partly described to the portion of uniqueness. The approximate 40 percent of the variance shared by the items seems to make the test sufficiently homogeneous to produce a reliable test, which differentiates students consistently according to proficiency of English.

The unexplained portion of the variance of the items is composed of the unique and error variance. In the framework of the classic test theory, it is possible to estimate the correlation between the scores obtained and the true scores by squaring the reliability estimates: .66, .61 of Group A, and .69, .69 of Group B for the first and second parts, respectively. By transforming the indexes into percentages, we can extract the common factors variances from them and obtain estimates of the unique variances of the variables. The analysis of the variances into separate components is displayed in Table 11.

Table 11. Variance components of items.

Subtests	Common factors		Uniqueness	Error	Total
	F I	F II, III			%
Items 1—10					
Written	31.8	7.2	27.0	34	100
Oral	35.1	5.8	20.1	39	100
Items 11—20					
Written	29.9	5.7	33.4	31	100
Oral	36.9	6.4	25.7	31	100

The patterns of loadings between the three factors were studied by computing the product moment correlation coefficient. Since the correlation coefficient gives an index of the patterns similarity, the coefficient of congruence was also computed to estimate an aspect of magnitude in similarity as well. The coefficient of congruence differs from the correlation coefficient in not standardizing the data. (Rummel 1970, 461).

Another technique of comparing the pattern—magnitude similarity of factor loadings is root mean square, which is more stringent than the coefficient of congruence. The root mean square is proportional to the Euclidean distance between factors. If the distance is zero, the two factors are similar in magnitude and direction. Departure from zero shows that the two factors are less alike. Since the root mean square coefficient measures any deviation between the two factors, it imposes very stringent similarity requirements on the comparison. (Rummel 1970, 461). All the three types of indexes are reported in Table 12.

The first factors of the communicative test seem to be similar in pattern and magnitude on the basis of the coefficients of congruence and root mean square. Items 11—20 produce a slightly larger deviation between the factors, obviously due to magnitude, because the correlation coefficient indicates an increased similarity in the patterns of the factors. The second and third factors do not show a consistent pattern similarity in the first and second parts of the test. Evidently the

factors have changed places in the two solutions, judged by the strength of congruence and correlation as well as by the change of positive and negative signs. There seems to be some similarity in magnitude between the factors but the pattern is inverse in the two solutions. This might be due to the different response modes. The aspects specific to the two factors are in connexion with each other, but in the inverse order. The similarity in magnitude and direction of the factor vectors differ to some extent as the root mean square coefficients indicate.

Table 12. Comparison between the factors solution for the written and oral response modes as indicated by coefficients of congruence, correlation and root mean square. (Decimal points omitted)

Factors	Coefficient of congruence			Correlation coefficient			Root mean square		
	I	II	III	I	II	III	I	II	III
Items 1—10									
I	99			60			07		
II		22			22			24	
III			-32			-32			27
Items 11—20									
I	98			66			12		
II		-51			-44			27	
III			22			23			20

Whether the amount of uniqueness, which was singled out from the total variance of the items, is characteristic of the communicative test, can be investigated by studying its empirical validity. This is made possible by the cloze and listening comprehension tests employed in the experiment. Correlations between them and the subtests of the communicative test were reported in Table 8. The written subtests correlated slightly more with cloze .69, .70 for groups A and B, respectively, than the oral subtests .65 and .67 for groups B and A, respectively. The correlations between the subtests and the listening comprehension test varied .64—.70.

Another set of criteria for comparison was obtained from the results of the matriculation examination, which the subjects took about four months after the data gathering of the experiment. The main results are reported in Table 13 and the intertest correlations in Table 14. Table 15 is made out to help the reader detect differences between correlations for the two groups.

Table 13. Results in the English matriculation examination of spring 1981 for the subjects of the experiment (N = 406) and for all the examinees (N = 28 732).

Variables No.	Maximum score	Groups				All population ¹			
		A		B					
		$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	s	r_{tt}	
5 Listening comprehension	90	71.4	12.1	73.2	11.6	69.8	12.9	.80	
6 Reading comprehension	90	71.3	11.9	72.3	11.6	69.8	13.3	.78	
7 Structures	18	12.2	3.1	12.1	3.2	11.6	3.4		
8 Essay	99	70.4	12.6	71.0	12.3	69.7	13.0		
9 Total score	297	226.9	39.4	229.0	33.5				
10 Final mark	6	4.2	1.3	4.2	1.3	4.0			

¹ Missing figures were not reported by the Examination Board.

Table 14. Intercorrelations of test variables for the groups separately. (Decimal points omitted.)

Variables	Group A										Group B									
	2	3	4	5	6	7	8	9	10		2	3	4	5	6	7	8	9	10	
1. Cloze	63	68	67	70	63	51	63	64	72		68	65	70	67	59	62	68	73	74	
2. Listening comprehension		67	63	63	69	54	50	78	75			67	72	70	63	52	59	72	73	
3. Items 1—10: Written/Oral			70	69	63	53	54	70	73				75	68	61	54	61	73	72	
4. Items 11—20: Oral/Written				68	56	44	60	57	66					63	63	55	62	72	70	
Matriculation examination:																				
5. Listening comprehension					73	45	59	73	84					71	51	59	85	85		
6. Reading comprehension						57	53	82	87						54	61	86	86		
7. Structures							53	67	67							63	67	65		
8. Essay								57	73									83	83	
9. Total score									92										96	
10. Final mark																				

Table 15. Differences of intercorrelations between Group A and Group B (A-B).

Variables	2	3	4	5	6	7	8	9	10
1. Cloze	-.5	.3	-.3	.3	.4	-.1	-.5	-.9	-.2
2. Listening comprehension		0	-.9	-.7	.6	.2	-.9	.6	.2
3. Items 1-10: Written/Oral			-.5	.1	.2	-.1	-.7	-.3	.1
4. Items 11-20: Oral/Written				.5	-.7	-.1	-.2	-.15	-.4
Matriculation examination:									
5. Listening comprehension						.2	-.6	0	-.13
6. Reading comprehension							.3	-.8	.4
7. Structures								-.10	0
8. Essay									-.26
9. Total score									-.10
10. Final mark									-.4

Correlations of the first part of the communicative test with the other tests show a similar pattern for the written and oral performances. The highest correlations are to be found among the second part of the communicative test (.70, .75), the total score (.70, .73) and the final mark (.73, .72). The lowest correlations are found among the subtests of structures (.53, .54), essay (.54, .61) and reading comprehension (.63, .61). In the second half of the communicative test, the lowest correlations are found among the same subtests. An exceptionally low correlation (.57) occurs between the oral performance and the total score of the matriculation examination, in comparison with .72 for the written performance. Another striking feature among the correlations between the written and oral performances is that structures seem to correlate even less with the oral performance (.44) than with the written (.55), which in itself might be considered natural but which was not found in the first half of the communicative test.

Listening comprehension is a component of communicative performance, but the listening comprehension tests do not show any very high correlations with either the oral or the written performance. Neither can the open response mode show a strong relation with the essay test as might be supposed because of the open-ended questions.

The correlations in themselves did not give a broad basis for inferences about the uniqueness of communicative performance in English. Therefore, a factorization with a two-factor solution under the principal axes model was performed on the tests for the groups separately. The total score and final mark variables were left out from the analysis because they were dependent on the subtests of the matriculation examination. The total score was a sum variable, and the final mark was given on the basis of the subtest scores. The numerical description is given in Table 16.

Table 16. The principal axes solutions of factors with the varimax rotation for the language tests, for the groups separately.

Tests	Principal axes				Varimax solution						
	A		B		A			B			h ²
	I	II	I	II	I	II	h ²	I	II	h ²	
1. Cloze	.82	.11	.83	.11	.70	.45	.69	.57	.61	.69	
2. Listening comprehension	.80	-.11	.82	-.17	.54	.59	.65	.75	.38	.70	
3. Items 1—10 Written/Oral	.82	.08	.82	-.14	.68	.47	.68	.73	.41	.70	
4. Items 11—20 Oral/Written	.80	.26	.84	-.11	.78	.31	.70	.72	.45	.71	
Matriculation examination:											
5. Listening comprehension	.83	.13	.81	-.16	.72	.44	.71	.74	.39	.69	
6. Reading comprehension	.81	-.20	.78	-.08	.49	.67	.69	.66	.43	.62	
7. Structures	.66	-.34	.71	.35	.29	.68	.55	.33	.71	.62	
8. Essay	.72	.01	.78	.25	.56	.46	.52	.45	.61	.68	
Eigenvalue	4.93	.27	5.12	.29	3.01	2.18	5.19	3.21	2.20	5.40	
Percentage variance	61.6	3.8	64.0	3.6	37.6	27.3	64.9	40.1	27.5	67.5	

The first factor for Group A explained 61.6 percent, and 63.9 percent for Group B of the total variance of the variables. The portions of the second factor were 3.3 and 3.6 for groups A and B, respectively. All the variables correlate with the factor .80 or slightly more, with the exception of the structures and essay variables (.66, .72 for A, and .71, .78 for B, respectively). Although the factors were rotated into the varimax criterion, no simple structure was to be found, because the variance was distributed on both the factors, giving no key to interpretation.

The pattern and structures of the factor solutions were investigated by computing the coefficients of congruence, correlation, and root mean square. Table 17 gives the indexes.

Table 17. Comparison between the factor solutions for groups A and B, as indicated by the coefficients of congruence, correlation, and root mean square. (Decimal points omitted.)

Factors	Coefficient of congruence		Correlation coefficient		Root mean square	
	I	II	I	II	I	II
I	1.0		88		03	
II		21		24		35

The first factor remains the same in the two solutions. Instead, the second factor seems to gather the loadings differently, and the vectors do deviate from each other to a considerable degree. Although a second factor can be singled out through factoring, it does not seem to be of the same aspect of language proficiency. It cannot, however, be attributed to the two response modes because its major contribution comes from the structures variable and the total score variable.

A comparison between the variance components reported in Table 11 and the amount of the variance explained by the first factors suggest that the uniqueness observed previously reduces markedly when other tests of language proficiency are included in the analysis. If there is something unique to be found with the use of a communicative test, the portion cannot be much more than a few percent of the total variance. This is, however, to say nothing of the importance of the value of the communicative test as a measuring instrument, which will be dealt with in the discussion further on.

8.7.3. The third problem area

The third problem area concerned the background factors which might have contributed to the performance of the communicative test. The analysis was based on the questionnaire completed by the subjects of the experiment. Table 18 gives the descriptive statistics of the variables classified according to the scheme displayed in Figure 7, Section 8.3.4.

The girls outnumbered the boys in the experiment, but the proportion of 57 % vs. 43 % is approximately the same as in the target population (61 % vs. 39 %) (Tilastotiedotus 1980). Out of the subjects in the experiment 45 % studied the long course of mathematics as compared with 39.1 % in the target population (Tilastotiedotus 1980). Some variation occurred in the number of years of English, due to failures and an alternative curriculum design of the language programme. In schools 4 and 5, some students had started to study Swedish prior to English, which reduced their English studies by two to three years. This variable does not correlate significantly with performance, other than with performance in the written subtest of the first part. The course chosen in mathematics has a statistically significant relationship with all the subtests. The negative sign is due to the manner of coding: 1 for the long course and 2 for the short.

Generally speaking, rather few students had attended language courses abroad or stayed a longer time in an English-speaking country. A few times situations had occurred in which students had spoken English to foreigners in Finland. Radio broadcasts had attracted them now and then. Obviously students took an interest in English books, newspapers and magazines, although it was not more frequent than once a month. Otherwise, they had not had many other activities which might have improved their English.

Some of the motivational inputs have marginally significant correlations with performance among one of the groups, but no consistent relationship occurs. An interesting variable is commitment to English studies. English seems to require as much work as any other academic subject. The statistically significant correlation with the performance of Group A tells us that those who work more are, however, poor in English. Apparently, they would not have kept up with the level they were at if they had not worked hard. The marginally significant correlation (.14) of the number of years of English also indicates difficulties in this direction.

The average attitude to the English teacher as inspiring, efficient, and pleasant was markedly positive among students. A possible source of error might be that students did not dare to express negative feelings. An attempt to control this was made by assuring students that the information on the questionnaire was confidential and that the answers

Table 18. Descriptive statistics of the variables of the questionnaire. Groups A and B, $n = 203$ in each.

Variables	A		B		Correlation with performance			
					A		B	
					Written	Oral	Oral	Written
	$\bar{X}$	s	$\bar{X}$	s	I	I	II	II
BACKGROUND								
1. Sex (girls 232, boys 174)	1.6	.5	1.6	.5	-.04	-.02	-.01	-.02
2. Amount of mathematics: long course 181, short 225	1.5	.5	1.6	.5	-.22	-.20	-.20	-.20
3. Years of English studies	9.5	1.4	9.5	1.4	.14	.12	.11	.02
MOTIVATION								
Student input in English								
4. English courses abroad	1.2	.7	1.2	.7	.23	.08	.21	.02
5. Stay abroad except courses	1.2	.8	1.2	.6	.25	.12	.20	.17
6. Communication with foreigners in Finland	2.3	1.0	2.3	1.0	.21	.11	.26	.11
7. English radio broadcasts	2.7	1.0	2.4	1.0	.07	.14	.17	.17
8. English books, papers, magazines	1.8	.9	1.7	.8	.18	.16	.23	.28
9. Other activities to improve English	.4	.6	.4	.6	.08	.18	.11	.24
22. Commitment to English studies	3.2	.9	3.1	.9	-.23	-.10	-.26	-.04
Teacher and his behaviour								
23. Inspiring	3.5	.9	3.4	.9	.01	.08	.11	.07
24. Efficient	3.8	1.0	3.7	1.1	.07	-.03	.10	.03
25. Pleasant	3.8	.8	3.8	.8	-.01	.01	.13	-.02

Variables	Correlation with performance							
	A		B		A		B	
					Written		Oral	
	$\bar{X}$	s	$\bar{X}$	s	I	I	II	II
ATTITUDE								
Attitude to instructional aims								
15. Pleasantness of oral exercises	3.6	1.0	3.6	.9	.15	.27	.29	.30
18. Pleasantness of written exercises	3.5	.9	3.5	.9	.04	.08	.03	.07
16. Easiness of oral exercises	2.7	1.0	2.9	.9	.31	.33	.37	.33
19. Easiness of written exercises	2.9	.9	2.9	.9	.25	.35	.24	.35
17. Importance of oral exercises	4.3	.7	4.3	.8	-.01	.06	-.03	.11
20. Importance of written exercises	3.5	.9	3.6	.8	-.05	-.21	-.10	-.18
21. Pleasantness of English studies	3.7	.9	3.5	.8	.02	.27	.11	.25
Attitude to the communicative test								
26. Easiness of oral response	2.1	1.0	2.6	1.1	.25	.42	.40	.36
27. Easiness of written response	3.0	1.1	2.7	1.0	.59	.24	.37	.40
28. Pleasantness of oral response	2.8	1.1	3.0	.9	.13	.40	.32	.24
29. Pleasantness of written response	3.4	.8	3.2	.9	.37	.16	.23	.26
METALANGUAGE								
Technical arrangements of instruction								
10. Visits to language laboratory	2.8	.7	2.8	.7	-.17	-.18	-.11	-.22
11. LC — speaking exercises in the laboratory	1.3	.6	1.3	.6	-.05	.10	-.09	.07
12. English tapes and recordings	3.3	1.1	3.3	1.1	.05	.10	.07	.07
Use of foreign/native language								
13. Teacher	3.9	1.0	3.7	1.0	-.02	-.11	-.07	-.09
14. Students	4.2	1.0	4.2	1.0	-.01	-.04	-.07	-.11

would not be shown to the teacher. Also, there is a possibility that students did not want to express negative feelings about their teacher to an outsider. On the other hand, one could wonder why a positive attitude should not exist among students. Nevertheless, the teacher and his behaviour do not correlate significantly with performance on the communicative test.

A similar positive attitude was shown to oral and written exercises as well as to English studies on the whole. However, students seemed to think that both written and oral exercises were fairly difficult. It was considered more important to practise oral than written skills although the latter were not all disregarded. Here the issue is not perhaps so much of preference for one or the other but rather, oral skills may not have been practised enough in class because of the written final examination.

Though written exercises were regarded as pleasant, it did not have any relationship with performance, either written or oral. Instead, pleasantness of oral practice correlates significantly (.27—.30) with written and oral performance, but only .15 (significant at 5 % level) with the written subtest of the first part. Easiness of oral and written exercises correlate positively with performance. The importance of written exercises has a significant negative relationship with performance of Group B. Those who performed well would have liked to reduce written exercises whereas those who did not do so well would have liked to increase them. This might be explained by the wish of poor performers to practise written work for the coming school-leaving examination.

The communicative test was considered rather difficult. Oral performance was felt to be slightly more difficult, especially by Group A. The *t* value is not, however, statistically significant. Also, the written version of the test pleased students slightly more than the oral one. The differences are not significant, though. The easiness and pleasantness of the communicative test have statistically significant correlations, and some of them even moderate, with performance. The relationship between the written subtest in Group A and the pleasantness of the oral version, however, is an exception (.13).

It seems to have become a custom in schools to attend the language laboratory once a fortnight as well as then to focus on listening comprehension exercises. Also, the use of English tapes and recordings in class are habitual practice, even though not in every lesson. The prevailing language in instructional interaction is English but the native language is obviously employed as an aid.

The variables under the component of metalanguage do not seem to affect performance in English. Only marginally significant, but negative correlations occur between the communicative test and visits to the language laboratory. The relationship is not significant for the oral performance of Group A. This can hardly be interpreted so that the fewer the visits to the laboratory the better the performance. Several

Table 19. Stepwise regression analysis with the communicative written performance as criterion and the background variables as predictors. (n = 203)

Variables	Written subtest of items 1-10										F	Significance
	Beta coefficients			t values			Increment in m c s					
	1.	2.	3.	1.	2.	3.	1.	2.	3.			
15. Mathematics	-.22	-.22	-.23	3.2	3.2	3.2	.05	.05	.05	10.4	.01	
16. Years of English		.18	.17		1.9	1.9		.02	.02	7.1	.01	
14. Sex			.06			.5			.00	4.8	.03	
Constant term	29.0	21.2	20.4									
R ²	.05	.07	.07									

explanations are available. For example, students might have liked more frequent visits to the laboratory. In their frustration they had a subjectively distorted feeling that the real attendance was less frequent than it actually was. Consequently, they expressed it by choosing either *a* or *b* from the alternatives. This view may be supported by what teachers told the researcher about their visiting the laboratory once a fortnight. On the other hand, the negative relationship is likely to be due to the inattentiveness of students to consider the past three years, and not only the last few months in answering the question. It goes without saying that laboratory exercises give way to other forms of exercises with the approaching matriculation examination in sight.

The capability of the three background variables to account for the communicative performance was investigated through regression analysis. The stepwise method was employed to see in what order the variables would enter the model. Only the outcome for the written subtest of the first part is reported in Table 19 to give the frame of the analyses. In the rest of the analyses the order of entries remained the same. The amount of mathematics alone was able to account for performance significantly ($R^2 .05$). This is also seen from the squared correlation coefficient.

The amount of mathematics alone explains 5 percent of the written performance. In the second run, the number of years of English offered an increment of two percent. The *t* value for beta is nonsignificant, though at a marginal level of 5 %. Sex has no predictive power at all. Contrary to the common belief that girls outdo boys in verbal tasks, they perform equally well on the communicative test.

The questionnaire was built on variables under the specific components. Sum variables were formed according to the original components, with such modifications that the background component was dropped and the subparts of the components were treated as separate variables (see Figure 7, p. 102). The means and standard deviations are reported in Table 20. The sum variables were correlated with the outcome of the communicative test. Table 21 shows the result.

Table 20. Means and standard deviations for the sum variables of the questionnaire by group.

Variables	A		B	
	$\bar{X}$	s	$\bar{X}$	s
101. Student input	12.8	2.9	12.3	2.8
102. Teacher and his behaviour	11.1	2.2	10.9	2.2
103. Attitude to instructional aims	24.1	2.8	24.2	3.0
104. Attitude to the communicative test	11.3	2.6	11.5	2.7
105. Technical arrangement of instruction	7.4	1.5	7.3	1.6
106. Use of foreign/native language	8.1	1.7	7.9	1.7

Table 21. Correlations between the outcome of the communicative test and the sum variables as well as intercorrelations of the sum variables by group.

	101	102	103	104	105	106
Written _A	22	03	24	51	-06	-02
Oral _B	21	02	34	46	03	-09
Written _B	26	03	35	47	-02	-12
Oral _A	27	14	31	51	-03	-08
A/B						
101		06	34	28	00	-10
102	13		-02	11	32	33
103	38	-08		53	07	01
104	31	14	38		-02	04
105	07	24	20	10		25
106	-04	25	18	-05	30	

The highest correlations are found between the attitude to the communicative test and performance. Second best was the variable of the attitudes to the instructional aims. Variable 101 also has a statistically significant relationship with proficiency. An attempt was made to explain English proficiency with the sum variables through stepwise regression analysis. Only the prediction of variable 104 was statistically significant. It was not even important to the instructional model. The increments of all other variables to explanation were one percent or less.

Correlations among variables varied to some extent, which suggested that they might cluster into factors, and thus offer a basis for explanation. An iterative analysis of 4 to 6 factors was performed under the principal axes model with estimates of the communality in the leading diagonal. The rotation programme used was the varimax procedure, which gave an orthogonal six-factor solution with the explaining portion of 31.5 percent of the total variance. To be on the cautious side, the arbitrary criterion of $\pm .30$ for the significance of factor loadings was employed for interpretation. A detailed description of the factor analysis is reported in Appendix 10. Table 22 gives the best loadings for the factors to be interpreted.

The backbone of the first factor consists of variables concerning the teacher and his instruction. A spurious variable of the pleasantness of written practice can perhaps be seen as an act of instruction. The first factor also loads on sex, giving the impression that girls tend to take an affective attitude to the teacher and instruction whereas boys do not.

Factors II and IV are, in the main, based on the attitude to oral proficiency and performance. The two factors are separated in the analysis but share something because they correlate .13 at a marginal level (.14 for 5 % significance). The five-factor solution did not provide a

Table 22. The six-factor varimax solution with the highest loadings for interpretation.

Variable	Loading	Factor I <i>Teacher and his instruction</i>
25	.73	Teacher pleasant
23	.72	Teacher inspiring
24	.53	Teacher efficient
21	.40	English studies pleasant
18	.38	Written practice pleasant
1	.32	Sex
12	.30	Listening to English tapes and recordings in class
Eigenvalue	2.3	
Percentage variance	8.0	
16	.65	Factor II <i>Liking for oral proficiency</i>
26	.58	Oral practice easy
28	.36	Oral test easy
18	-.34	Oral test pleasant
20	-.32	Written practice pleasant
		Written practice be reduced
Eigenvalue	1.5	
Percentage variance	5.3	
2	.62	Factor III
10	.57	Mathematics (short course)
7	.30	Frequency of visits to language laboratory
		Listening to English broadcasts
Eigenvalue	1.2	
Percentage variance	4.2	
15	.62	Factor IV <i>Liking for oral performance</i>
28	.51	Oral practice pleasant
6	.50	Oral test pleasant
17	.42	Speaking English to foreigners in Finland
8	.40	Oral practice be increased
		Frequency of reading English books, papers or magazines
7	.31	Staying in an English-speaking country
Eigenvalue	1.9	
Percentage variance	6.4	
29	.51	Factor V <i>Liking for written proficiency and performance</i>
27	.51	Written test pleasant
19	.49	Written test easy
18	.40	Written practice easy
		Written practice pleasant
Eigenvalue	1.2	
Percentage variance	4.2	
13	.54	Factor VI <i>Metalanguage</i>
12	.37	Teacher spoken English in class
11	.32	Listening to English tapes and recordings in class
14	.32	Practice of oral skills in language laboratory
		Students spoken English in class
Eigenvalue	1.0	
Percentage variance	3.4	

clearer interpretation. The distinguishing feature is perhaps that factor II represents a more cognitive attitude to oral proficiency whereas factor IV centres round an affective attitude to oral language, which manifests itself in an active communicative role.

Factor III is difficult to interpret. It mainly loads on mathematics, frequency of visits to the language laboratory, and marginally on listening to English broadcasts. The factor is connected with instructional arrangement, giving the interpretation that those who study the short course of mathematics tend to visit the language laboratory more frequently. This perhaps implies an instructionally diversified practice in English teaching.

Factor V loads on the written test and written practice, justifying the name of written proficiency and performance for the factor. Factor VI refers rather clearly to the component of metalanguage, including oral exercises in the laboratory too.

The interrelations of the factors were computed by means of factor score variables. Table 23 displays them.

Table 23. Intercorrelations of factor score variables (N = 406). (Decimal points omitted.)

Factors	I	II	III	IV	V
II	-04				
III	10	04			
IV	04	13	05		
V	06	04	-04	12	
VI	09	02	00	-05	02

As mentioned above, the highest but nonsignificant correlation is between factors II and IV. The fourth factor also correlates somewhat with factor V (.12). Otherwise, the relationships are negligible, thus confirming the orthogonality of the solution.

The computation was continued to investigate through regression analysis how well the factors could explain the oral and written performances of the communicative test. Since the factors did not correlate with each other because of the orthogonal rotation, the squared correlations provide the same information as the standardized beta coefficients. Also, on the basis of the factor solution, it might have been possible to form sum variables and to investigate how well those variables would be able to predict performance. The procedure would have been very unlikely to produce more information than the factor scores because the original correlations between the performance and the variables of the questionnaire were not very high. As a consequence, only the correlation matrix for the performance and the factors is displayed in Table 24.

Table 24. Intercorrelations between performance on the communicative test and the six factors. (Decimal points omitted.)

Subtests	I	II	III	IV	V	VI
Written _A	-02	27	-23	22	49	04
Oral _B	07	39	-20	31	31	03
Written _B	03	34	-19	32	45	01
Oral _A	09	40	-17	34	33	-00

Factor V can best predict performance on the communicative test, particularly written proficiency as the name of the factor implies. Factors II and IV also have the power to predict, oral proficiency slightly better than written proficiency this time. As expected on the basis of correlations, the first and the sixth factor could not predict performance in the communicative test.

As an aspect of test consistency, transferability is the agreement of the test with itself across varieties of people (Cattell 1978, 282). The stability of the factors in the form of emerging in another sample was studied by factoring the variables for the individual groups. The comparison of the factor solutions was made by means of the coefficients of congruence, correlation, and root mean square. There are certain limitations and risks in these indexes, which Cattell (1978) discusses widely. He even suggests a nonparametric *s* for comparison. It is not, however, unbiased, either. Here we content ourselves with the three indexes to obtain a rough estimate of similarity. The detailed outcomes are given in Appendix 11.

The two six-factor varimax solutions did not quite agree in similarity in the two groups. Factors emerged in a different order in the varimax solutions, and partly distributed differently on variables. Despite that, some congruence coefficients are fairly high. Since the sampling distributions of congruence coefficients and root mean square are not known, tests of significance of matches are precluded.

The final question about the questionnaire is its reliability: the agreement of the test with itself across two samples of the same population. The outcome for the matched groups separately showed an approximate equality of means and standard deviations. This gives proof of equal variances. The similar trends between scores could be estimated by correlations, which serve as the basis of reliability estimates. An alpha coefficient would not be a correct index, because it measures internal consistency on the basis of correlations between items and the sum variable over all items for subjects. The computation method is subject-centred. It is justified when the reliability of achievement tests is estimated. An individual's scoring high, in regard to the other subjects and other items, is not relevant to an attitudinal scale.

Rather, the stability of scores might be estimated by a parallel test, i.e., the test is given to another sample of the population. By adding scores of one group for a particular item enables forming two sum variables. The correlation between them could be considered an index of reliability. In this study, the homogeneity of the items of the questionnaire, i.e. the common factors, were employed as the basis for the estimate. It can be considered an item-based approach, which focuses on the measuring instrument, instead of on the individuals' performance. The separate factor solutions for the groups were utilized to compute the correlation between the communalities. The index obtained was .84, which manifests high stability and reliability. For comparison, an alpha coefficient was also computed with the result of .70. The index would thus have given a distorted picture of reliability.

8.8. Discussion of the experiment

In answer to the first problem, we can state that this experiment did not succeed in rejecting the hypothesis of the unified structure of proficiency in a foreign language. Since the statistical test of no difference was not significant, the number of subjects was relatively large, and no obvious causes are known to violate the results, we can conclude that a difference between the response modes is also very unlikely to occur in the population. Our confidence in the outcome is confirmed by the control of the various levels of verbal ability. Part of the confidence can be contributed to the experimental design, though it was conducted under field circumstances. It is not possible to achieve complete control of the laboratory conditions, but one way of approaching such a state is to preclude all potential sources that might render vulnerable the results. One weakness lies in the repetition of the testing events. An attempt was made to standardize them as much as possible. On the other hand, schools did not compete with each other, but the experimental design as a whole was repeated in the different settings. This is likely to level out the discrepancies between the settings, because comparison was not made between them in making inferences. Thus, the potentially different climates of schools and classes could influence the experiment just as any element of the population in a random sample.

Some positive support for the hypothesis of the unitary structure of language proficiency was obtained from the factor analysis of the different language tests. More than sixty percent of the variance loaded on a single factor. However, the factoring did not produce a simple structure. Therefore, conclusions must be drawn with caution. The portions of the second factors were small, only 3.3 — 3.6 percent of the total variance of the variables. The characteristics of the second factors

seem to be an aspect of written proficiency, obviously related to the object language. The aspect may be a result of a narrow scope of language proficiency, probably having something to do with vocabulary and structures.

What was left over from the common variance after the first factor had been extracted, loaded on all variables. In this regard, the communicative test did not differ from the other tests. On the whole, the uniting feature for the tests on which the first factor most strongly loads is that the tests measure integrated language proficiency. What strikes one is the lower loadings of the structures and essay tests. If these tests of the matriculation examination are reliable (no indexes available), they measure aspects, other than are usually manifest in the tests of foreign language proficiency. The communicative test shared 29—38 percent of the total variance with the essay, and only 19—30 percent of the variance with the structures. From this we can conclude that the communicative test does not require quite the same qualities as essay writing, although open-ended answers might suggest that.

The outcome of the experiment implies two questions. First, the intonation and pronunciation were ignored in testing the hypothesis of the unified structure of language proficiency. Second, if the communicative test does not yield any other information than what is obtained through other language tests, we can question its utility, in regard to the previous ones.

The first question may be answered by reference to the experimental design. The two response modes were compared to test the hypothesis of no difference. Since the oral answers had to be evaluated by means of listening to them, the comprehensibility of intonation and pronunciation were actually indirectly measured. Otherwise, the performance could not have been assessed. If the intonation and pronunciation had been taken into account, it would have led to the assessment of their correctness, naturalness, possibly beauty, etc., in regard to the native speakers' performance. Thus, another aspect for evaluation would have been included, but it would hardly have yielded any substantial information in terms of communication.

Since understanding the spoken answers and their grammatical correctness were evaluated, the essential communicative features of the linguistic code were involved. Consequently, the omission of intonation and pronunciation is unlikely to distort the outcome of the experiment. Of course, this is not to underrate the importance of good pronunciation and intonation. They are certainly factors which contribute to the total performance level of foreign language proficiency, in terms of fluency and acceptance among the native speakers of the language. The argument is only for the result of the experiment.

For obvious reasons, this experiment cannot show whether the aspect of good intonation and pronunciation is a separate skill, or part of the

unified structure of language proficiency. The first alternative has obtained some support from empirical findings in which the skill has emerged as a separate factor (Löfgren 1972a, 33). On the whole, this experiment cannot give any *new* information on the issue of the unified structure, because the yield mainly corroborates the previous findings.

The problem of specific factors remains open, because their contribution is probably so slight. They emerge in factorization but it is difficult to specify them. As a matter of fact, the approach of factorizing different language tests can be made questionable as a method of solving the problem. Tests differ in so many aspects that we can never know for sure in interpretation what the variables on which a factor loads actually have in common. The construction of tests, the way of questioning, what is questioned, the response manners, etc. cause immense variation. Since these things may awaken different human processes, possibly require different skills, it must be an overwhelming problem to decide whether a process, skill or a particular linguistic element is involved in a specific factor which does not contribute much to the total variance explained by the variables. One way of tackling the problem might be a more intensified approach in which focus is centred on measuring a particular process or skill at a time. Then testing may be widened to concern various contributing aspects. This makes the different tests more commensurable and possibly enables inferences on features, common to the variables of a factor.

Judging from the small amount of variance explained by the factors of items, it seems reasonable to assume that there were other elements of information included in the items. The estimates of uniqueness by means of reliability indexes indicates this. Traditionally, items are viewed as a sample of the content domain. That would be a sample of the object language. But a variety of skills, knowledge, psychological factors, etc. are also involved in testing. Since these phenomena are not inherent parts of the items but of the students, they may emerge in the unique portion of the variance. Since the yield of the communicative test in the form of common and unique variance was estimated slightly greater than the common variance it shared with the other language tests, we may hypothesize that the extra portion embraces factors specific to discourse processing. The written version, it is true, included more uniqueness. More investigation is then needed into its character to find out whether it is irrelevant of spoken response.

Though the items shared only about 40 percent of the common variance, the communicative test proved reliable, judged from various angles. The coefficient alpha, as an index of the internal consistency to differentiate students, was good, though the subtests had only ten items each. The reliability clearly rose when the two subtests were combined. The good reliability is evidently mainly because students produced answers and did not choose alternatives. This asset might have been lost

in the scorer's subjectivity but was obviously avoided by limiting the answer range. A certain amount of subjectivity has been found in global evaluation, for example in assessing essays. To reduce subjectivity, analytical evaluation has been resorted to. In marking the answers of the communicative test, the advantages of both methods were combined. The questions were based on short passages so that the answers became restricted in range. This helped the student to remember better what he heard, and also, it made the answers more comparable. This, in addition to the background information of the text shared by the student and the scorer, facilitated the scorer's task to set up criteria for assessment. In spite of the measures taken to control subjectivity, some of it always remains in the test scores. Whether it is in constructing the test or assessing the responses, or in both, requires being conscious of it and attempting to control it as far as possible.

The facility or difficulty of a test is perhaps more important to students than to evaluators. Although there exist no absolute criteria, the intuitive percent could be at least fifty. The facility value in the second part of the communicative test remained below that level. The average for the matriculation examination seems to be 70 percent, and even higher for multiple-choice tests. For productive tasks it is perhaps too high a percent. The lower values in the communicative test may be due to the inexperience of the students with this kind of test. In absolute terms, the material was not too difficult. This was also confirmed by the teachers. All vocabulary and structures should have been known to the students. One thing that may have made the test more difficult than anticipated was the slight background noise on tape, because the test had not been recorded in a studio. The inclusion of only ten items per test might be risky if the test is too difficult. Then the performance of weaker students will be based on too few items. The use of twenty items may be recommended.

Although no difference was detected between the response modes, the question arises whether the oral and written responses were quite similar. Thus, a short discussion is justified. In oral answering, students repeated, corrected words and expressions, coughed and sometimes murmured. Some pauses for thinking and extra sounds and hesitations occurred, though, as a rule, the response was given promptly. Students' concentration on answering was often intense, and sometimes a weak sigh of relief indicated the end of the performance. Frustration in the inability to answer was reflected as sighing and puffing, mistakes as short laughs. Some students gave natural exclamations, such as: Sorry, I was distracted for a moment, I didn't expect such a question, etc. Moreover, this shows that it is risky to have too few items in a test, because casual distraction and inattentiveness may cause inability to answer. The spontaneous reactions were not revealed in written answers. It is natural because forty seconds is a short time for writing. Such

spontaneity can give a more vivid picture of a student's fluency and character, but whether it adds to the information on his language proficiency is not certain.

One might suppose that written answers are linguistically more correct, but the experiment did not prove it. In answering orally, the student monitors his speech all the time and tends to correct his errors. The reason why written and oral correctness was equal may partly be due to the short answers, which enabled monitoring correctness with the limits of memory. This can also be a flaw in the communicative test. In fact, the question is a complex one, because many casual, psychological and situational factors may affect performance both in real and testing situations, and people do differ in this respect, so that it is not certain whether the phenomenon is a real shortcoming for the test.

In addition to reliability, validity is an important aspect for the development of a test. Reliability is a necessary condition for validity. The preceding discussion of the experiment has given empirical evidence for the theoretical development of the previous parts of this study, which thus constitutes the construct validity of the communicative test. The term 'content validity' is used to refer to the relevance and representativeness of the test items as a sample of a specified behavioural domain about which inferences are to be drawn. As Messick (1981) states this has more to do with test construction than test validity. The communicative test as a measurement of discourse processing cannot be confined to a particular content domain of behaviours. Rather, it concentrates on a single process: processing discourse, which, in fact, has two facets. The content validity is not revealed in particular items but as a whole in the subject-matter and the linguistic material, which must be in conformity with the educational aims. As long as the questions asked of students concern the subject-matter (of the text), we can say that the relevant process is tested. If it seems relevant action to teachers and students, which there is no reason to doubt, the test can be considered to have so-called face validity, too.

Concrete evidence of the students' attitudes to the communicative test was obtained from the questionnaire. The average difficulty was slight and the pleasantness somewhat above the mean score. The written response mode was considered slightly easier and it was somewhat favoured, though the differences were not significant. The pleasantness of the written response mode is obviously due to more familiarity and practice with it. The sum variable of the students' attitudes to the communicative test was the best predictor of performance.

The purpose of the questionnaire was to control some background variables which might have distorted the results. The issue concerns the generalizability of the outcome of the experiment. The second purpose was to explore some tentative connexions between communicative proficiency in English and some factors within the framework of the

instructional model.

An interesting finding was that the sex did not affect performance in the communicative test, although girls are generally believed to excel over boys in verbal tasks. Neither did the number of years of English studies affect performance. This must not be judged in absolute terms. It is obvious that the time devoted to studies plays a part, but one or two years' difference at the beginning of many years' studies does not apparently make a distinction in proficiency. The differences are likely to level out in the senior secondary school, particularly, because those with a shorter course behind them are mixed with more advanced students.

The choice of a long or short course in mathematics had a statistically significant connexion with English proficiency. It explained five percent of the performance. Those who studied the long course performed slightly better. Whether it is due to talent or to instruction and practice with a formal language, or to both, cannot be answered by this research. One might hypothesize on the basis of some previous evidence that talent for mathematics and languages often go hand in hand.

The motivational factors measured as the students' input in English did not emerge in great frequency. Language courses and stays abroad were not very common, but they have a statistically significant relation with performance. The strength of relationship must here be interpreted tentatively because of the restriction of range in computing the correlations. Since these out-of-class activities, however, have a positive, though not strong, correlation with performance, it is obviously advisable to steer motivation in that direction. On the other hand, the outcome may be a consequence, rather than a cause, even though more practice is likely to improve proficiency.

The variable of commitment to English studies does not give an explicit interpretation. It was thought to be evidence of a person's wish and interest to devote her time to English. However, its negative connexion with performance may indicate a must to work more with English. The average point (3.2) is slightly above the mean and perhaps reflects the difficulty of English for a majority of students. Also, the direct questions of the difficulty of exercises confirm this view. It might imply that more time for instruction is needed, or the requirements are too demanding within the resources of the course and the whole curriculum.

The students' attitudes to the teacher and his behaviour were investigated as a source of motivation for proficiency. They were markedly positive, and the teacher was experienced as inspiring, efficient and pleasant. Nevertheless, this has no direct connexion with performance. This finding is contrary to common belief. As a matter of fact, teachers may find the fact discouraging. It goes without saying that studying in pleasant surroundings and atmosphere is more agreeable.

The effect may be indirectly reflected in the students' devotion to English as a lifelong interest. The pleasantness of the instructional environment is bound to promote the growth of students in the direction of broader educational goals.

The students' attitudes to English instruction and to the importance and pleasantness of it are markedly positive, despite its difficulty. The importance of oral exercises was emphasized more than that of written exercises. The difficulty also has a statistically significant connexion with proficiency. The students have a realistic view of themselves. If they find studies easy, they also perform better than those who find them difficult.

The pleasantness of English studies has not quite clear-cut relations with performance. If oral exercises were considered pleasant, students also performed better, both in the written and oral versions of the communicative test. If written exercises were found pleasant, it did not have influence on performance. We might suppose that, on the whole, a positive attitude to oral skills reflects a more global approach to language learning than if only written exercises are found pleasant. The outcome will be better performance if the aims of the language learning are seen broadly.

Generally speaking, the importance of oral or written exercises does not seem to have a significant connexion with proficiency. A marginally significant correlation shows rather a negative connexion. Thus the result from the three factors of pleasantness, difficulty, and importance is slightly contrary to some previous studies, which have concluded that all three undeniably contribute to performance in academic subjects.

The component under the heading of metalanguage gave interesting information about the arrangement of instruction. Though English was a prevailing vehicle of communication in instruction, it did not seem to have an effect on performance. If it did, it was negative, rather than positive. The same connexion with performance applies to the visits to the language laboratory and the exercises there. This area needs more research. If the finding is valid, it suggests that the language laboratory is not an important vehicle of instruction in terms of learning listening comprehension. Perhaps less technical apparatus is enough, such as tapes and tape-recorders. The connexion of laboratory exercises with performance is not statistically significant, either, but however, positive. The finding is somewhat surprising and needs further research. A larger proportion of oral exercises might have yielded another outcome. A possible flaw is that the result of the metalanguage component is based on the students' subjective experience of the frequency of the arrangements, and not on factual information. On the other hand, the means are rather high, and the standard deviations are not very small, except for the language laboratory variables, so that a technical error might be discarded.

The data of the questionnaire did not provide us with information

that might distort the results of the experiment. The proportions between girls and boys, and between long and short courses of mathematics did not show marked deviations from those of the target population. Also, the importance of out-of-class activities in the form of language courses or long-term stays abroad was of no significance to the experiment. Since the use of language laboratories did not affect the learning outcomes, we may generalize even to the target population, although the frame population was made up of schools which were equipped with at least 15 student tape-recorders in their language laboratories.

On the whole, connexions between performance and the factors within the framework of the instructional model remained scarce and tentative. A specific and more detailed investigation into relevant variables is needed to obtain evidence for a further development of the instructional model. A general instructional model may be a more difficult goal to achieve than one within foreign language pedagogy. ▼

The communicative test was treated above as if it had constituted a unified entirety. It was due to the similarity and equality of written and oral performance on the communicative test of oral proficiency in English. No one study, or science in general, is able to give the final word, but nevertheless, some generalizations and conclusions must and can be made. Even recommendations are required, particularly in education, about the use of the test.

The communicative test as such seems to be a reliable and valid instrument of measuring discourse processing. One would like to add, in a foreign language, though this testing was confined to English. Logically and intuitively, there is no reason to believe that the results would not apply to any foreign language. Further research may find empirical evidence for this. This study focused only on the question — answer type of testing, but the vehicle may be considered a relevant instrument of measuring communicative proficiency in a foreign language. The asset is its simplicity and administrability, even to large numbers of testees in fairly reduced, simple circumstances.

For our matriculation examination, an even more simplified version could be recommended, because the written version to replace the oral responses proved to be an equally reliable and valid measuring instrument. One drawback is then that intonation and pronunciation are ignored, and might even be neglected in instruction. This should be prevented by teachers, whose responsibility it is to have them practised. Listening comprehension exercises would also lead to indirect practice of pronunciation and intonation. Besides, it is very unlikely that reaction fluency is achieved without practising conversation and discussion in the foreign language.

The advantages of the written version of the communicative test are obviously greater in terms of economy, reliability and administrability.

Equipping every school with language laboratories where tape-recorders are installed in every student's booth is too expensive. Administering a simultaneous test for more than 30 000 students by recording their answers, evaluating them, and the like, makes the task unrealistic. Then the written version could offer a more economical, administrable and realistic alternative, which however is relevant in terms of communicative oral proficiency in a foreign language. As a productive test, it would overcome the drawbacks of the multiple-choice tests, and however, attempt to measure oral proficiency. The time that is spent on preparing good multiple-choice tests might be invested in the scoring of written answers to the questions. It would take more time, and the subjectivity of scoring is increased, but it can be greatly reduced with appropriate measures.

An important factor in preparing an instrument is the formulation of questions so that good answers may be expected. Questions which induce yes-no answers should be avoided. The students in the experiment did not show a tendency to content themselves with one-word answers. Besides, testees will gradually learn the strategy with practice. Although it is impossible to anticipate all possible answers, decisions on acceptance can be agreed upon.

As for the administration, the present circumstances in schools are sufficient to carry out the test, either in the language laboratory or in the normal classroom with the additional loudspeakers. Experience has shown that, the more technical apparatus and arrangements are needed, the more risks are taken of having disturbances when it comes to all senior secondary schools.

9. Conclusion of the Research

This study is based on four broad problem areas which form a hierarchy from a broad scope to a concentrated focus. The treatment of the problems proceeds from the theoretical level to the level of practical realization. The problems are tackled in a logic order so that the outcome of solving the first leads to solving the second, the outcome of which then leads to the third, etc. The results do not give mutually redundant information, but the constituent parts needed for solving a problem are partly overlapping in the different problems. Each problem area has an internal structure of its own, which can be viewed as a whole. Theoretically, it consists of a hierarchy whose lower levels underlie the higher levels.

Empirical research is usually steered by the theoretical framework and previous studies. Inferences are based on statistical outcomes which are accepted with certain criteria for probabilities. Therefore, this always entails further research. The empirical content for the evaluation model was aimed at by the experiment. The interest was confined to communicative oral proficiency. The experiment was also built on the hypothesis of the unified structure of foreign language proficiency. The hypothesis could not be rejected, so that it thus obtained more corroboration. It was also found that a slight portion of the structure of language proficiency might be specific to a particular skill, but the finding can still be considered tentative. The hints may serve as an inspiration for further investigation.

One of the purposes of the experiment was to explore a new type of test on the basis of the evaluation model. Since it proved good and useful in many respects, it could be recommended for use in testing. Naturally, there are weaknesses in the test which need further clarifications. They cannot, however, be obtained without experience.

The significance of the functional model for evaluation depends on the validity of the theoretical background. Correspondence with reality does not make the model rational but rather how good its conceptual definition is. This is the core of the study, because the evaluation model is subordinated to the instructional curriculum. The operationalization of it is described as a model for discourse processing, which serves as a method in a broad sense. Evaluation is connected with this method through the functional model. The logic tenability of this construction is the power and weakness of this research, because the deductions and inferences were made on the basis of it. Further research and criticism will show whether the construction is meaningful and useful. We consider it to clarify and facilitate the development of subject-related curricula and to serve as the framework for making syllabuses, analyzing language proficiency, and constructing tests. A further development is

needed so that learning strategies can be related to the framework as a meaningful and functional part. Specifying the model and making it more precise on logical grounds and on the basis of empirical evidence may lead to an instructional theory in foreign language pedagogy.

The justification for the development of the construction can be seen as a problem of validity. Communicative language proficiency was instructionally operationalized into discourse processing. Communicative language proficiency, as a term of ordinary language, was analyzed and defined in this study. Thus it has obtained the position of a scientific concept. The name has two facets. First, it distinguishes the concept from the corresponding linguistic concept of communicative competence, which was seen to serve another purpose. Second, the connexion with everyday language is preserved to mark the relationship so that no new term would falsely lead to the conception that something else is meant. Since foreign language pedagogy involves practical action, it is important that the chasm between theoreticians and practitioners is not widened. In this study, communicative proficiency in a foreign language is an analyzed, scientific concept, which can be interpreted as a conceptual operationalization of the communicative proficiency in a foreign language.

The intensional meaning of the concept was analyzed on the basis of the literature related to the field of study. The term 'communicative competence' was disclosed as a creation and development of theoretical linguistics and its applied fields. Since research within those fields of study is done at an abstract level of language, and since this is investigated in terms of the native language, the pragmatic approach to language analysis was resorted to. The basis of selection for literature was the principle that the conceptions, irrelevant of foreign language learning and instruction, could be eliminated to achieve a concrete concept of communicative proficiency in a foreign language. This was done by the attempt to find the features that are transmitted in the linguistic code, and the extralinguistic factors which are not necessary for learning and teaching a foreign language. For that reason, the communicative meaning of the linguistic code was analyzed as reflected in the discourse analysis researched by language philosophers and pragmatic linguists.

An emphasis on and an analysis of the linguistic material in this study is exceptional for educational research. The reason for this is that traditionally, linguistic research has had a strong influence on foreign language teaching. The purpose of the present study was to stress the importance of investigating instruction within foreign language pedagogy, a sub-area of education. Therefore, it was necessary to display the difference between the concepts of linguistics and education. As soon as the concept of communicative competence became widespread and its origin obscure, it was not clear what it actually meant as a

learning aim. For that reason, the concept had somehow to be operationalized. This research can be viewed as a hierarchical operationalizing process of the concept from the perspective of foreign language pedagogy.

References

- Abercrombie, D. 1973. Paralinguistic communication. In Allen, J. P.B. & S. P. Corder (eds) *Edinburgh course in applied linguistics 1*. London: Oxford University Press, 31—36.
- AILA 1976. Proceedings of the fourth international congress of applied linguistics. I—III. Ed. by G. Nickel. Stuttgart: HochschulVerlag.
- Alikoski, J. 1969. Käytännön otantaongelmista kasvatustieteessä ja niiden vaikutuksesta tutkimuksen suunnitteluun. *Kasvatusopillinen aikakauskirja* CVI, 190—195.
- Austin, J.L. 1962. *How to do things with words*. London: Oxford University Press.
- Ausubel, D.P. & F.G. Robinson 1969. *School learning*. Bristol: Holt, Rinehart and Winston.
- Bacon, W. & S-L. Ojanen 1976. Foreign language conversation testing at Tampere Language Institute. In Kohonen, V. & L. Nummenmaa (eds) *Special issue on testing communicative competence. Kielikeskustutisia 4*, 55—59.
- Berk, R.A. (ed.) 1980. *Criterion-referenced measurement: the state of the art*. Baltimore & London: The John Hopkins University Press.
- Berlo, D.K. 1977. Communication as process: review and commentary. In B.D. Ruben (ed.) *Communication yearbook 1*. New Brunswick, New Jersey: Transaction Books, 11—27.
- Blankertz, H. 1975. *Theorien und Modelle der Didaktik*. (9., neuarbeitete und erweiterte Auflage. München: Juventa Verlag.
- Bloom, B.S. 1971. Learning for mastery. Ch. 3 in Bloom, B.S., J.Th. Hastings & G.F. Madaus *Handbook of formative and summative evaluation of student learning*. New York: McGraw-Hill.
- Bloom, B.S., J.Th. Hastings & G.F. Madaus 1971. *Handbook of formative and summative evaluation of student learning*. New York: McGraw-Hill.
- Brière, E. 1980. Communicative competence, variable rules, and interdisciplinary research. In Scarcella, R.C. & S.D. Krashen (eds) *Research in second language acquisition*, 89—94.
- Campbell, R. & R. Wales 1970. The study of language acquisition. In Lyons, J. (ed.) *New horizons in linguistics*. Harmondsworth: Penguin.
- Carroll, J.B. 1961. Fundamental considerations in testing for English language proficiency. Conference on testing the English proficiency of foreign students. Washington, D.C.: Center for Applied Linguistics, 30—40. Reprinted in Allen, H.B. & R.N. Campbell (eds) *Teaching English as a second language*. New York: McGraw-Hill, 1972, 313—320.
- Carroll, J.B. 1963a. Linguistic relativity, contrastive linguistics, and language learning. *IRAL*, 1—20.
- Carroll, J.B. 1963b. A model for school learning. *Teachers College Record* 64, 723—733.
- Cattell, R.B. 1978. *The scientific use of factor analysis in behavioral and life sciences*. New York: Plenum Press.
- Child, D. 1970. *The essentials of factor analysis*. London: Holt, Rinehart and Winston.
- Chomsky, N. 1965. *Aspects of the theory of syntax*. Cambridge, Massachusetts: The M.I.T. Press.

- Clark, J. 1972. Foreign-language testing: theory and practice. Philadelphia: The Center for Curriculum Development.
- Clark, J.L.D. 1975. Theoretical and technical considerations in oral proficiency testing. In Jones, R.L. & B. Spolsky (eds) Testing language proficiency. Center for Applied Linguistics. Arlington, Virginia, 10—28.
- Cohen, A. 1975. Measuring intelligibility in foreign-language learning. In van Essen, A.J. & J.P. Munting (eds) The context of foreign-language learning. Assen, the Netherlands: van Gorcum & Comp. B.V., 5—17.
- Cook, Th. D. & D.T. Campbell 1979. Quasi-experimentation. Design and analysis issues for field settings. Chicago: Rand McNally College Publishing Company.
- Cooper, D.E. 1973. Philosophy and the nature of language. London: Longman.
- Corder, S.P. 1973a. Introducing applied linguistics. Harmondsworth: Penguin.
- Corder, S.P. 1973b. Linguistic theory and applied linguistics. In Corder, S.P. & E. Roulet (eds) Theoretical linguistic models in applied linguistics. (3rd AIMAV Seminar Neuchâtel 1972). Bruxelles: AIMAV.
- Coulthard, M. 1977. An introduction to discourse analysis. London: Longman.
- Criper, C. & H.G. Widdowson 1975. Sociolinguistics and language teaching. In Allen, J.P.B. & S.P. Corder (eds) Edinburgh course in applied linguistics 2. London: Oxford University Press, 155—217.
- Cronbach, L.J. 1951. Coefficient alpha and the internal structure of tests. *Psychometrika* 16, 297—334.
- Darnell, D. 1968. The development of an English language proficiency test of foreign students using a clozentropy procedure. ERIC ED 024—039.
- Davies, A. (ed.) 1968. Language testing symposium. London: Oxford University Press.
- Davies, A. 1971. Aptitude for and proficiency in French in the first year of the U.K. secondary school. In Perren, G.E. & J.L.M. Trim (eds) Selected papers of the second international congress of applied linguistics, Cambridge 1969. Cambridge University Press.
- Derwing, B.L. 1973. Transformational grammar as a theory of language acquisition. London: Cambridge University Press.
- Diller, K.C. 1976. Criteria for adapting language teaching methods to the learning styles and abilities of students. *AILA*, 341—351.
- Ebel, R.L. 1972. Essentials of educational measurement. 2nd ed. Englewood Cliffs, New Jersey: Prentice-Hall.
- Edwards, A.L. 1968. Experimental design in psychological research. New York: Holt, Rinehart & Winston.
- van Ek, J.A. 1976. The threshold level for modern language learning in schools. Council for Europe, DECS/EGT (76) 19.
- Ellegård, A. 1977a. Hur lär vi oss språk? 1. *Moderna språk* 71, 3, 207—220.
- Ellegård, A. 1977b. Hur lär vi oss språk? 2. *Moderna språk* 71, 4, 313—324.
- Enkvist, N. & V. Kohonen 1976. Cloze-testien teoriaa. In Kohonen, V. (ed.) Integratiivisen kielitaidon mitaamisesta cloze-testien avulla: teoriaa ja sovelluksia. Turku: AFinLA, No. 13.
- Folland, D. 1976. Group conversation as a means of teaching and testing spoken language. In Kohonen, V. & L. Nummenmaa (eds) Special issue on testing communicative competence. *Kielikeskustutistia* 4, 45—53.
- Galisson, R. & D. Coste 1976. Dictionnaire de didactique des langues. Paris: Librairie Hachette.
- Galloway, L. & S.D. Krashen 1980. Cerebral organization in bilingualism and second language. In Scarcella, R.C. & S.D. Krashen (eds) Second language

- acquisition. Rowley, Mass.: Newbury House, 74—80.
- Germain, C. 1973. La notion de situation en linguistique. Ottawa: Editions de l'université d'Ottawa.
- Green, S.B., R.W. Lissitz & S.A. Mulaik 1977. Limitations of coefficient alpha as an index of test unidimensionality. *Educational and Psychological Measurement* 37, 827—838.
- Grice, H.P. 1975. Logic and conversation. In Cole, P. & J. Morgan (eds) *Syntax and semantics* 3. New York: Academic Press, 41—58.
- Habermas, J. 1971. Vorbereitende Bemerkungen zu einer Theorie der kommunikativen Kompetenz. In Habermas, J. & N. Luhmann *Theorie der Gesellschaft oder Sozialtechnologie*. Frankfurt: Suhrkamp.
- Halliday, M.A.K. 1973. *Explorations in the functions of language*. London: Edward Arnold.
- Harris, D. 1969. *Testing English as a second language*. New York: McGraw-Hill.
- Hays, W. 1963. *Statistics*. London: Holt, Rinehart & Winston.
- Heikkurinen, T. 1971. Tiedostamistavan muutokset ja opettamisen vaikuttamislunne. Tutkimuksia n:o 14. Kasvatustieteen laitos. Helsingin yliopisto.
- Heinonen, V. & E. Viljanen 1980. *Evaluatio koulussa*. Helsinki: Otava.
- Hellgren, P. 1976. Cloze-testi suomenkielisen lukion oppilaiden englannin kielen taidon mittarina. *Kasvatustieteen laudaturtyö*. Helsingin yliopisto. (Unpublished.)
- Hellgren, P. 1980. Vieraskielen kommunikatiivinen taito ja sen arvioiminen. *Kasvatustieteen lisensiaatintutkimus*. Kasvatustieteen laitos. Helsingin yliopisto. (Unpublished.)
- Hendrickx, J. 1977. Developing listening comprehension materials and tests. In Dirven, R. (ed.) *Hörverständnis im Fremdsprachenunterricht*. Listening comprehension in foreign language teaching. Regensburg: Scriptor Verlag.
- Hirvonen, P. 1971a. Englannin kielen taidon mittaaminen lukion päättyessä 1: Lahtokohdat. Turku: AFinLA, N:o 1.
- Hirvonen, P. 1971b. Englannin kielen taidon mittaaminen lukion päättyessä 2: Kuuntelukoe. Turku: AFinLA, N:o 4.
- Hirvonen, P. 1973. Englannin kielen taidon mittaaminen lukion päättyessä 4: Puhumieskoe. Turku: AFinLA, N:o 8.
- Howatt, A. 1974. The background to course design. In Allen, J.P.B. & S. P. Corder (eds) *Techniques in applied linguistics* 3. London: Oxford University Press.
- Hymes, D.H. 1972. On communicative competence. In Pride, J.B. & J. Holmes (eds) *Sociolinguistics*. Aylesbury, Bucks: Penguin.
- Ingram, E. 1968. Attainment and diagnostic testing. In Davies, A. (ed.) *Language testing symposium*. London: Oxford University Press, 70—97.
- Irvine, P., P. Atai & J.W. Oller Jr. 1974. Cloze, dictation, and the test of English as a foreign language. *Language Learning* 24, 2, 245—252.
- Itkonen, E. 1974. Linguistics and metascience. Kokemäki: Societas Philosophica et Phaenomenologica Finlandiae. *Studia Philosophica Turkuensia* Fasc II.
- Jakobovits, L.A. 1970. Foreign language learning. A psycholinguistic analysis of the issues. Rowley, Mass.: Newbury House.
- Johansson, S. 1973. An evaluation of the noise test - a method for testing overall second language proficiency by perception under masking noise. *IRAL* XI, 2, 107—133.
- Jöreskog, K.G. 1979. Statistical models and methods for analysis of longitudinal data. In Jöreskog, K.G. & D. Sörbom *Advances in factor analysis and*

- structural equation models. Ed. by J. Magidson. Cambridge, Mass.: Abt Books, 129—169.
- Karlsson, F. 1977. Johdatusta yleiseen kielitieteeseen. Helsinki: Gaudeamus.
- Kielikello 10. Kielitoimiston tiedotuslehti. Kotimaisten kielten tutkimuskeskus.
- Kohonen, V. 1976. Cloze-testien ongelmia testianalyysin valossa. In Kohonen, V. (ed.) Integratiivisen kielitaidon mittaamisesta cloze-testien avulla: teoriaa ja sovelluksia. Turku: AFinLA, No. 13.
- Komiteamietintö 1973: 38. Oppilasarvostelun uudistamistoimikunnan mietintö. Helsinki 19.
- Kontinen, R. 1980. Testiteoria. Johdatus kasvatus- ja käyttäytymistieteellisen mittauksen teoriaan. Helsinki: Gaudeamus.
- Koort, P. 1974. Edukation och planering. Helsingfors: Gaudeamus.
- Koskenniemi, M. & K. Hälinen 1970. Didaktiikka. Helsinki: Otava.
- Kostera, P. 1979. Vieraiden kielten puhumiskokeesta. *Tempus* 4, 18—24.
- Krashen, S. 1977. The monitor model for adult second language performance. In Burt, M., H. Dulay, & M. Finocchiaro (eds) Viewpoints on English as a second language. New York: Regents.
- Lado, R. 1961. Language testing. London: Longman.
- Lahdes, E. 1977. Peruskoulun uusi opetusoppi. Helsinki: Otava.
- Lahti, O. 1977. Monivalintaa on liikaa. *Tempus* 9, 27—31.
- Leeson, R. 1975. Fluency and language teaching. London: Longman.
- Lehtonen, J. 1972. Kielenopetuksen fonologiaa. Kasvatustieteiden tutkimuslaitoksen julkaisuja 168. Jyväskylän yliopisto.
- Lehtonen, J., K. Sajavaara & A. May 1977. Spoken English. The perception and production of English on a Finnish-English contrastive basis. Jyväskylä: Gaudeamus.
- Leimu, K. & H. Saari 1976. Koulukoe- ja arviointitoiminnan kehittämisestä Suomessa. Selosteita ja tiedotteita 77. Kasvatustieteiden tutkimuslaitos. Jyväskylän yliopisto.
- Leino, A.-L. 1972. English school achievements and some student characteristics I. Research Bulletin No. 33. Institute of Education. University of Helsinki.
- Leino, A.-L. 1978. An introduction to the content of foreign language instruction. Research Bulletin No. 50. Institute of Education. University of Helsinki.
- Leino, A.-L. 1979. Kielididaktiikkaa. Helsinki: Otava.
- Leino, J. 1977. Matematiikan didaktiikka I. Helsinki: Kirjayhtymä.
- Lindquist, E.F. 1953. Design and analysis of experiments in psychology and education. Boston: Houghton Mifflin Co.
- Lord, F.M. & M.R. Novick 1968. Statistical theories of mental test scores. London: Addison-Wesley.
- Low, O. 1979. Certificate of proficiency English practice. London: Edward Arnold.
- Lusensky, B. 1972. Suggestions for a listening comprehension test. Engelska institutionen. Pedagogiska institutionen. Göteborgs universitet. Rapport 8. Projekt SPRENG.
- Lyons, J. 1963. Structural semantics. Oxford: Blackwell.
- Lyons, J. 1977. Semantics. I—II. Cambridge: Cambridge University Press.
- Löfgren, H. 1969. Mätningar av språkfärdighet i tyska: en undersökning på elever i årskurs 7. Pedagogisk-psykologiska problem. Nr. 84. Pedagogisk-psykologiska institutionen. Lärarhögskolan, Malmö.
- Löfgren, H. 1972a. Mätningar av språkfärdighet i tyska: en undersökning på elever i årskurs 9. Pedagogisk-psykologiska problem Nr. 158. Pedagogisk-

- psykologiska institutionen. Lärarhögskolan, Malmö.
- Lotfren, H. 1972b. Mätningar av språkfärdighet: En hypotesprövande undersökning av en faktorstruktur. Pedagogisk-psykologiska problem Nr. 161. Pedagogisk-psykologiska institutionen. Lärarhögskolan, Malmö.
- Mackey, W.F. 1965. Language teaching analysis. London: Oxford University Press.
- Mackey, W.F. 1973. Language didactics and applied linguistics. In Oller, J.W. Jr. & J. Richards (eds) Focus on the learner: pragmatic perspectives for the language teacher. Rowley, Mass: Newbury House, 4—13.
- Malinen, P. 1974. Kasvatustieteen metodologia. Opetusmoniste 43. Kasvatustieteen laitos. Jyväskylän yliopisto.
- Malmberg, B. 1966. Structural linguistics and human communication. Berlin: Springer-Verlag.
- Malmberg, B. 1971. Applications of linguistics. In Perren, G.E. & J.L.M. Trim (eds) Selected papers of the second international congress of applied linguistics. Cambridge University Press.
- Malmberg, B. 1973. Språkinläring. Stockholm: Bokförlaget Aldus/Bonniers.
- Melenk, H. 1977. Der didaktische Begriff der "kommunikativen Kompetenz". *PRAXIS des neusprachlichen Unterrichts* 1, 3—12.
- Messick, S. 1981. Evidence and ethics in the evaluation of tests. *Educational Researcher* 10, 9, 9—20.
- Miller, G.A. & S. Isard 1963. Some perceptual consequences of linguistic rules. *Journal of Verbal Learning and Verbal Behavior* 2, 217—228.
- Mindt, D. 1978. Probleme des pragmalinguistischen Ansatzes in der Fremdsprachendidaktik. *Die Neueren Sprachen* 3/4, 340—356.
- Munby, J. 1978. Communicative syllabus design. London: Cambridge University Press.
- Nunnally, J.C. 1967. Psychometric theory. New York: McGraw-Hill.
- Nykykielet 1971. Valtion oppikoulujen opetussuunnitelmat. Helsinki: Valtion painatuskeskus.
- Oakeshott-Taylor, J. 1977. Information, redundancy, and listening comprehension. In Dirven, R. (ed.) Hörverständnis im Fremdsprachenunterricht. Listening comprehension in foreign language teaching. Regensburg: Scriptor Verlag, 93—112.
- Oller, J.W. Jr. 1971. Dictation as a device for testing foreign-language proficiency. *English Language Teaching* 25, 254—259.
- Oller, J.W. Jr. 1973a. Discrete-point tests versus tests of integrative skills. In Oller, J.W. Jr. & J. Richards (eds) Focus on the learner: pragmatic perspectives for the language teacher. Rowley, Mass: Newbury House, 184—199.
- Oller, J.W. Jr. 1973b. Pragmatic language testing. *Language Sciences*, 7—12.
- Oller, J.W. Jr. 1973c. Cloze tests of second language proficiency and what they measure. *Language Learning* 23, 1, 105—118.
- Oller, J.W. Jr. & Ch. A. Conrad 1971. The cloze technique and ESL proficiency. *Language Learning* 21, 2, 183—195.
- Oller, J.W. Jr. & F.B. Hinofotis 1980. Two mutually exclusive hypotheses about second language ability: indivisible or partially divisible competence. In Oller, J.W. Jr. & K. Perkins (eds) Research in language testing. Rowley, Mass.: Newbury House, 13—23.
- Pasanen, U-M. 1977. Päätötkokeen uudistuksen heijastuminen lukion vieraankielen opetukseen opettajien arvioimina. Research Reports No. 57. Department of Education. University of Jyväskylä.

- Pearson, A. T. 1980. The competency concept. *Educational Studies*. A journal in the foundations of education 11, 2, 145—152.
- Perren, G.E. 1968. Testing spoken language. Some unsolved problems. In Davies, A. (ed.) *Language testing symposium*. London: Oxford University Press, 107—116.
- Piepho, H-E. 1974a. Kommunikative Kompetenz, Pragmalinguistik und Ansätze zur Neubesinnung in der Lernzielbestimmung im Fremdsprachenunterricht. Düsseldorf.
- Piepho, H-E. 1974b. Kommunikative Kompetenz als übergeordnetes Lernziel im Englischunterricht. Dornburg—Frickhofen.
- Popham, W.J. 1978. *Criterion-referenced measurement*. Englewood Cliffs, N.J.: Prentice Hall.
- Poyatos, F. 1976. Verbal and nonverbal expression in interaction: research and pedagogical perspectives. *AILA* I, 87—97.
- von Raffler-Engel, W. 1976. Some rules of socio-kinesics. *AILA* II, 113—121.
- Riley, Ph. 1976. Current trends in language teaching at the C.R.A.P.E.L. In Kohonen, V. & L. Nummenmaa (eds) *Special issue on teaching and testing communicative competence. Kielikeskusuutisia* 4, 9—25.
- Rivers, W. 1968. *Teaching foreign-language skills*. Chicago: The University of Chicago Press.
- Rummel, R.J. 1970. *Applied factor analysis*. Evanston: Northwestern University Press.
- Sajavaara, K. & J. Lehtonen 1977. Finnish-English contrastive project, Occasional Papers No. 1. Suomen ja englannin kontrastiivinen projekti: tavoitteet ja menetelmät. Jyväskylä.
- Scholz, G., D. Hendricks, R. Spurling, M. Johnson & L. Vandenburg 1980. Is language ability divisible or unitary? A factor analysis of 22 English language proficiency tests. In Oller, J.W. Jr. & K. Perkins (eds) *Research in language testing*. Rowley, Mass.: Newbury House, 24—33.
- Searle, J.R. 1969. *Speech acts*. Cambridge: Cambridge University Press.
- Seeleye, H.N. 1968. *Teaching culture strategies for foreign language teachers*. Skokie, Ill.: National Textbook Co. 148.
- Slama-Cazacu, T. 1971. *Langage et contexte. Le problème du langage dans la conception de l'expression et de l'interprétation par des organisations contextuelles*. La Haye: Mouton.
- Slama-Cazacu, T. 1976. Applied psycholinguistics: its objects and goals. *AILA* I, 27—64.
- Spolsky, B. 1976. Language testing: art or science. *AILA* III, 9—28.
- Spolsky, B., B. Sigurd, M. Sato, E. Walker & C. Arterburn 1968. Preliminary studies in the development of techniques for testing overall second language proficiency. *Language Learning*, Special Issue 3, 79—101.
- Stampe, D. 1975. Meaning and truth in the theory of speech acts. In Cole, P. & J.L. Morgan (eds) *Syntax and semantics* 3. New York: Academic Press, 1—39.
- Stendahl, Ch. 1970. Testing English as a second language at university level. G. The Speaking Test. Rapport 16. Engelska & Pedagogiska institutionen. Göteborgs universitet.
- Strawson, P. 1964. Intention and convention in speech acts. *Philosophical Review* 73, 439—460.
- Sörbom, D. 1979. Detection of correlated errors in longitudinal data. In Jöreskog, K.G. & D. Sörbom *Advances in factor analysis and structural equation models*. Ed. by J. Magidson. Cambridge, Mass.: Abt Books,

- 171—184.
- Taylor, W. L. 1953. Cloze procedure: a new tool for measuring readability. *Journalism Quarterly* XXX, 415—433.
- Trim, J.L.M. 1971. Towards a situational definition of language contents in foreign languages. Council of Europe, CCC/EES 135.
- Trim, J.L.M. 1977. Report on some possible lines of development of an overall structure for a European unity credit scheme for foreign-language learning by adults. Strasbourg: Council for cultural co-operating council for Europe.
- Tilastotiedotus 5.12.1979. Tilastokeskus. No. KO 1979:15.
- Tilastotiedotus 19.12.1980. Tilastokeskus. No. KO 1980:16.
- Upshur, J.A. 1971. Productive communication testing: progress report. In Perren, G.E. & J.L.M. Trim (eds) Selected papers of the second international congress of applied linguistics, Cambridge 1969. Cambridge: Cambridge University Press, 435—441.
- Valette, R.M. 1967. Modern language testing. New York: Harcourt, Brace & World.
- Valette, R. 1971. Evaluation of learning in a second language. In Bloom, B.S., J. Th. Hastings & G.F. Madaus (eds) Handbook of formative and summative evaluation of student learning. New York: McGraw-Hill, 815-853.
- Walz, H.P. 1976. The human brain as the source of linguistic behaviour. *AILA* I, 99—107.
- Whitaker, H.A. 1975. Linguistics and neurology. In Bartsch, R. & Th. Vennemann (eds) Linguistics and neighboring disciplines. Holland: North-Holland Publishing Company, 33—42.
- Widdowson, H.G. 1978. Teaching language as communication. London: Oxford University Press.
- Wilds, C.O. 1975. The oral interview test. In Jones, R.L. & B. Spolsky (eds) Testing language proficiency. Arlington, Virginia: Center for Applied Linguistics, 29—44.
- Wittgenstein, L. 1953. Philosophische Untersuchungen — Philosophical investigations. Oxford: Basil Blackwell.
- Wilkins, D.A. 1976. Notional syllabuses. London: Oxford University Press.
- Williams, F. 1970. Language and poverty. New York: Academic Press.
- Winer, B.J. 1971. Statistical principles in experimental design. 2nd. ed. New York: McGraw-Hill.
- von Wright, G.H. 1968. Logiikka, filosofia ja kieli. Helsinki: Otava.
- Wörterbuch der Pädagogik in drei Bänden. 1977. Hrsg. vom Wilemann-Institut München—Wien. Freiburg: Herder.

Appendices

1. The transcript of the communicative test	160
2. The cloze test	164
3. The questionnaire in Finnish and English	165
4. The schools of the experiment marked on the map of Finland	170
5. A list of observations on the answers per questions	171
6. A diagram of the score dispersion in the communicative test as four subtests	175
7. Item analysis of the communicative test	176
8. Interitem correlations of the communicative test	177
9. A factor analysis for items of the communicative test	178
10. A factor analysis for the questionnaire	180
11. Coefficients of congruence, correlation and root mean square between the two factor solutions of the questionnaire for groups A and B separately	182

Appendix 1.

The transcribed text and the questions of the communicative test recorded on tape

I

Interviewer: Last summer I met Peter Bull from London in an international summer course in the German Democratic Republic. I asked him a few questions. Tell me, Peter, something about what you do for a living.

Peter: I'm a school teacher. I teach German in a large London comprehensive school. The school is situated in the southern part of London, about ten kilometres from the town centre, in Streatham.

I: Can you describe in a few words what your school is like?

P: Yes, it's a comprehensive school — has fourteen hundred pupils — is eight form entry, and I'm principally involved in teaching pupils from the second year onwards, namely, pupils of twelve to eighteen.

1. How old are Peter's pupils?

When I say comprehensive, it's a Catholic school as well, and we in England do understand the possibility of having a school which is not only comprehensive but Catholic, too, which may seem to be a bit of a contradiction.

2. What is special about Peter's school?

I: You say it is a Catholic school. Does it mean that the teachers and pupils have to be Catholic?

P: Yes, the pupils must certainly, and out of the fourteen hundred pupils that we have, they are all Catholic, and have to have a letter from their priest to say that their parents are practising members of the Catholic faith.

3. How must the pupils prove that they come from Catholic families?

The teachers are not all Catholics because it's very difficult sometimes to find Catholic teachers in certain specialist subjects. German is an example of that.

4. All the teachers are not Catholics in Peter's school. Why not?

I: What kind of school-leaving examination have you got?

P: Myself, I qualified with three A levels, that being the short for advanced level, and that's the examination you take when you are eighteen. We specialize to a considerable degree, and so, at A level, most pupils who are studying at A level will take three subjects — not necessarily English or maths, but a combination of subjects which they hope will provide them with suitable qualifications for further employment. Not as in some Continental schools, whereby you have to take the language of your country.

5. *What is said about English in the school-leaving examination?*

I: What, if any foreign languages are compulsory subjects?

P: Every pupil in our school learns French from the first year onwards, so from the eleventh year onwards they study French. And then pupils who've shown some ability at French would in their next year be offered a choice of doing German or Italian.

6. *What are the students like who start to learn German or Italian in Peter's school?*

I: As you teach German, your pupils are selected, but on the whole, have you got any discipline problems in your school?

P: As far as discipline is concerned, I think it would be true to say that as our school is a Catholic school, then the influence of the family, as well as the church is stronger than it might be in other state schools. And so for that reason, our pupils are perhaps slightly more disciplined, although I do get the impression that our school is perhaps noisier than others.

7. *Why do pupils in Peter's school behave, perhaps better than pupils in other state schools?*

I: I've heard that severe punishments, even corporal punishment, are sometimes used in English schools. Is that true?

P: Corporal punishment really means, I suppose, beating people with a stick, either on your behind or on the hand, is not indiscriminate, hitting at the pupils by teachers. And although this has been a tradition in English education, it's slowly dying out. And in our school, too, as from next February, corporal punishment will be abolished.

8. *What is the position of corporal punishment in English schools?*

I: Is there any improvement or a more democratic attitude in view in the near future?

P: Yes, curiously enough, I've been given a task just this term of trying to set up a pupils' council which would offer some vehicle for representation of pupils' views, which has only been possible on an individual level before, i.e. [ai i:] from pupil to teacher. And so we now have to set up a council which will be some representative body for public opinion within the school.

9. *What is done in Peter's school to increase democracy?*

I: Have you any idea about what problem pupils consider the biggest in schools?

P: Well, yes, I think it possibly has something to do with the pupil teacher relationship, whereby the distance which exists between pupils and teachers is resented by other pupils, and they feel that it would be more acceptable if the teachers had closer contact with the pupils in order to help motivate them, in order to help explain the various problems which arise during the course of their education.

10. *What is wrong with the pupil teacher relationship in Peter's school?*

I: Peter, you told me before that you studied at Oxford University. Is it easy to get into such an old university as Oxford?

P: Well, from the statistical point of view, it is not any more difficult to get into Oxford than it is to any other university in the country. The major difference lies perhaps in the fact that not so many schools apply to Oxford because they're rather frightened of the supposedly high academic standard.

11. Why don't all new comprehensive schools send their pupils to Oxford University?

And it is true that Oxford recruits principally still from the public schools and grammar schools to the extent whereby the newer comprehensive schools feel rather frightened of the competition.

12. What schools do most Oxford students come from?

I have the impression too, that academic standards are very high at Oxford. Nevertheless, that's not to say that those academic standards are not being met by the schools, the newer comprehensive schools, in our country.

13. What does Peter say about the standards of the comprehensive schools compared to those of Oxford?

I: What is it like to study in such an appreciated university with long traditions?

P: It is somewhat different from other universities in the sense that one has a personal tutor whom one sees every week or three times a fortnight, to whom one has to present an essay, which has been the subject of one's research for the previous week.

14. What do students at Oxford do to show that they have worked hard?

I: What activities did the students have besides their serious studies?

P: Oxford is famous for the wealth of its various societies, and also for its eating houses, and for its pubs. So, therefore, given also the tremendously cosmopolitan mix-up of people there, at Oxford, one could say that the possibilities for outside activities are really larger at Oxford than at any other university.

15. Apart from eating and drinking, what else can students do outside their studies in Oxford?

But on the other hand, of course, one could say that the college system makes people rather confined in the sense that they really don't go out to meet too many other people. It's almost a contradiction in a way.

16. After all, students are not very eager to go out. Why not?

I: I've heard you speak German very fluently. You can't possibly have learnt all your German in Oxford.

P: No, as a matter of fact, the course that we studied at Oxford was very academic, and so for that reason, although there were

opportunities, I didn't make use of them to practise my German in Oxford.

17. What did Peter think of the possibility to learn to speak German in Oxford?

I studied, well, when I say I studied, I spent one year as a language assistant in Germany. And I'd also been there beforehand, a couple of years before, as a guest pupil at a school.

18. How did Peter learn to speak German fluently?

I: Is the improvement of the language the only reason why you've now come to the German Democratic Republic?

P: No. Not the only reason at all, I also wanted to learn more about the country. Basically, because very little seems to be known about the German Democratic Republic in England, or rather more, not very much information is — is given out about the German Democratic Republic. And I find this rather awkward in my lessons, because I find myself teaching about Germany and meaning thereby the Federal Republic, which, of course, is really a misrepresentation of the truth.

19. What false picture has Peter so far given to his pupils about Germany?

I: In a way, this is not only a holiday for you but an intensive course. Do you find this a good means of spending your holiday?

P: Yes, indeed. Because on the one hand, I've been able to relax in the pleasant surroundings of Erfurt. But also, I've had intellectual stimuli from the various people around me, from the course offered. When I go back, I therefore have lots of new things to give to my pupils, I will have a broader view of what the concept Germany means, and above all, I will have had the pleasant personal experiences of meeting people from other countries.

20. What has made Peter's holiday a good one?

Appendix 2.

CLOZE TEST

Directions: (1) Read the whole passage through. (2) Then, go back and fill in the blanks with the words you think are missing. (3) Use only one word for each blank.

The other day, SOME two or three days AGO actually, I visited my SISTER in hospital. She is OLDER than I am: in FACT she was almost grown-up WHEN I was born. She HAD had a minor operation BUT was reported to have LARGELY got over the after-EFFECTS of this. My wife suggested MY taking her some flowers and FRUIT but she said that I NEED not take HER any books as she HERSELF had given her a LOT earlier. This was my FIRST visit so it took ME several minutes to locate THE ward she was IN. I always find hospital visiting SOMETHING of an ordeal, and AS I walked between the ROWS of beds, I felt VERY depressed, especially, at the SIGHT of the lonely patients, THOSE whose friends had not TURNED up to chat with THEM. My sister was half-sitting, half-LYING, propped up on pillows, WITH a paperback detective story IN her hand. She put ASIDE the book, a little RELUCTANTLY it seemed, thanked me IN a rather half-hearted way FOR the gifts and eyed ME silently while I took a SEAT on the hard WOODEN chair by her bedside. I commented on how WELL she looked but she IGNORED my remark. "Tell me," she BURST out unexpectedly, "if you WERE a doctor and wanted to COMMIT a murder, ONE of your patients, of course, how WOULD you do it?"

Appendix 3.

NIMI _____

etunimi sukunimi (tekstaten)

Vastaa kysymyksiin täydentämällä tai rengastamalla oikea vaihtoehto.

1. Olen a. poika b. tyttö
2. Lukiossa olen opiskellut matematiikkaa a. pitkän b. lyhyen kurssin.
3. Montako vuotta olet opiskellut englantia koulussa luokallejäämiesit mukaan lukien? _____
4. Oletko osallistunut englannin kielikursseihin ulkomailla? a. en b. kyllä Jos vastasit kyllä, kuinka pitkän ajan yhteensä viikoissa laskettuna? _____
5. Oletko oleskellut englanninkielisessä maassa muuten kuin kielikursseihin osallistumalla? a. en b. kyllä Jos vastasit kyllä, kuinka pitkän ajan viikoissa tai kuukausissa yhteen laskettuna? _____
6. Oletko Suomessa joutunut puhumaan englantia ulkomaalaisten kanssa? a. tuskin koskaan b. pari kertaa c. silloin tällöin d. melko usein e. erittäin usein
7. Oletko seurannut englanninkielisiä radiolähetyksiä? a. en b. pari kertaa c. silloin tällöin d. säännöllisesti 1-2 kertaa kuussa e. lähes viikottain
8. Luetko koulukirjojen lisäksi joko englanninkielisiä kirjoja tai lehtiä? a. tuskin koskaan b. yhden kuukaudessa c. yhden kahdessa viikossa d. yhden viikossa e. yhden joka päivä f. useampia? _____
9. Katsotko suullisen tai kirjallisen englannin taitosi parantuneen jollakin muulla koulun ulkopuolisella toiminnalla? Selitä lyhyesti.

Seuraavat kysymykset koskevat englannin opetusta lukion aikana:

10. Kielistudiossa olemme käyneet
a. tuskin koskaan b. kerran kuussa c. kerran kahdessa viikossa
d. kerran viikossa e. useammin kuin kerran viikossa
11. Studiokurssit on käytetty pääasiassa
a. puheen ymmärtämisen harjoitteluun lähes yksinomaan
b. puheen ymmärtämisen ja puhumisen harjoitteluun suhteessa 2:1
c. puheen ymmärtämisen ja puhumisen harjoitteluun suhteessa 1:1
d. puheen ymmärtämisen ja puhumisen harjoitteluun suhteessa 1:2
e. puhumisen harjoitteluun lähes yksinomaan
12. Tunneilla englanninkielisiä nauhoitteita (esim. kirjan kappale, uutisia, kieliradio-ohjelmia) on kuunneltu säännöllisesti
a. joka tunti b. joka toinen tunti c. silloin tällöin d. epäsäännöllisesti mutta harvoin e. tuskin koskaan
13. Englannin opettaja on puhunut oppitunneilla
a. lähes yksinomaan englantia b. englantia enemmän kuin suomea
c. yhtä paljon englantia kuin suomea d. vähemmän englantia kuin suomea
e. tuskin koskaan englantia
14. Oppilaat ovat puhuneet ja vastanneet englannin tunneilla
a. lähes yksinomaan englanniksi b. englanniksi enemmän kuin suomeksi
c. yhtä paljon englanniksi kuin suomeksi d. vähemmän englanniksi kuin suomeksi e. tuskin koskaan englanniksi

KÄÄNNÄ

Seuraavassa tiedustellaan mielipidettäsi englannin opiskelusta:

15. Suullisen kielitaidon harjoittelu on mielestäni
a. hyvin miellyttävää b. miellyttävää c. en osaa sanoa
d. epämiellyttävää e. hyvin epämiellyttävää
16. a. hyvin vaikeaa b. vaikeaa c. en osaa sanoa d. helppoa
e. hyvin helppoa
17. Suullisen kielitaidon harjoittelua on mielestäni
a. vähennettävä paljon b. vähennettävä hieman c. en osaa sanoa
d. lisättävä hieman e. lisättävä paljon
18. Kirjallisen kielitaidon harjoittelu on mielestäni
a. hyvin miellyttävää b. miellyttävää c. en osaa sanoa
d. epämiellyttävää e. hyvin epämiellyttävää
19. a. hyvin vaikeaa b. vaikeaa c. en osaa sanoa d. helppoa
e. hyvin helppoa
20. Kirjallisen kielitaidon harjoittelua on mielestäni
a. vähennettävä paljon b. vähennettävä hieman c. en osaa sanoa
d. lisättävä hieman e. lisättävä paljon
21. Englannin opiskelu koulussa on mielestäni
a. hyvin epämiellyttävää b. epämiellyttävää c. en osaa sanoa
d. miellyttävää e. hyvin miellyttävää
22. Muihin lukuaineisiin verrattuna työskentelen englannin parissa
a. paljon vähemmän b. vähemmän c. suunnilleen yhtä paljon
d. enemmän e. paljon enemmän
23. Englannin opettajani on mielestäni vaikuttanut opiskeluuni
a. hyvin innostavasti b. melko innostavasti c. en osaa sanoa
d. kielteisesti e. hyvin kielteisesti
24. Mielestäni englannissa meitä on opetettu
a. hyvin tehottomasti b. melko tehottomasti c. en osaa sanoa
d. melko tehokkaasti e. hyvin tehokkaasti
25. Englannin opettajani on mielestäni ollut
a. hyvin miellyttävä b. miellyttävä c. en osaa sanoa
d. epämiellyttävä e. hyvin epämiellyttävä
- Arvioi äskeisessä puheenymmärtämiskokeessa suullisen ja kirjallisen
vastaamisesi helppoutta/vaikeutta sekä miellyttävyyttä/epämiellyttävyyttä.
Mielestäni oli
26. Suullinen: a. hyvin helppoa b. melko helppoa c. en osaa sanoa
d. melko vaikeaa e. hyvin vaikeaa
27. Kirjallinen: a. hyvin helppoa b. melko helppoa c. en osaa sanoa
d. melko vaikeaa e. hyvin vaikeaa
28. Suullinen: a. hyvin miellyttävää b. miellyttävää c. en osaa sanoa
d. epämiellyttävää e. hyvin epämiellyttävää
29. Kirjallinen: a. hyvin miellyttävää b. miellyttävää c. en osaa sanoa
d. epämiellyttävää e. hyvin epämiellyttävää

KIITOKSIA PALJON VAIVANNÄÜSTÄSI !

NAME

first name family name (in block letters)

Answer the questions by completing them or by circling the appropriate alternative.

1. I am a. a boy b. a girl
2. At the senior level I have studied a. a long course b. a short course of mathematics.
3. How many years have you studied English at school, with failure years included?
4. Have you attended English language courses abroad? a. No b. yes
If you answered *yes*, for how long in total, counted in weeks?
5. Have you stayed in an English-speaking country *other* than attending language courses? a. No b. Yes
If you answered *yes*, for how long in total, counted in weeks or months?
6. Have you spoken English with foreigners in Finland?
a. hardly ever b. a couple of times c. occasionally d. fairly often e. very often
7. Have you listened to radio transmissions in English? a. no b. a couple of times c. occasionally d. regularly once or twice a month e. almost weekly
8. In addition to the school books, do you read *either* English books or magazines? a. hardly ever b. one a month c. one a fortnight d. one a week e. one every day f. more than one?
9. Do you find your oral or written skills of English have been improved by some other out-of-school activity? Describe briefly.

The following questions concern the instruction of English at the *senior secondary* level.

10. We have attended the language laboratory
a. hardly ever b. once a month c. once a fortnight d. once a week e. more frequently than once a week
11. The lessons in the language laboratory have been mainly used for
a. listening comprehension exercises almost exclusively
b. listening comprehension and speaking exercises in proportion 2:1
c. listening comprehension and speaking exercises in proportion 1:1
d. listening comprehension and speaking exercises in proportion 1:2
e. speaking exercises almost exclusively
12. In class English recordings (e.g. the text of the lesson, news, radio language programmes) have been regularly listened to
a. every lesson b. every two lessons c. occasionally d. irregularly but seldom e. hardly ever

13. In class your English teacher has spoken.
a. English almost exclusively b. English more than Finnish (Swedish) c. English as much as Finnish (Swedish) d. English less than Finnish (Swedish) e. English hardly ever
14. In English class students have spoken and answered questions
a. in English almost exclusively b. in English more than in Finnish (Swedish) c. in English as much as in Finnish (Swedish) d. in English less than in Finnish (Swedish) e. in English hardly ever

In the following questions you are asked to give your opinion on English studies:

15. In my opinion practising *oral skills* is
a. very pleasant b. pleasant c. don't know d. unpleasant e. very unpleasant
 16. a. very difficult b. difficult c. don't know d. easy e. very easy
 17. In my opinion practising oral skills will have to be
a. greatly reduced b. slightly reduced c. don't know d. slightly increased e. greatly increased
 18. In my opinion practising *written skills* is
a. very pleasant b. pleasant c. don't know d. unpleasant e. very unpleasant
 19. a. very difficult b. difficult c. don't know d. easy e. very easy
 20. In my opinion practising written skills will have to be
a. greatly reduced b. slightly reduced c. don't know d. slightly increased e. greatly increased
 21. In my opinion studying English at school is
a. very unpleasant b. unpleasant c. don't know d. pleasant e. very pleasant
 22. Compared with other academic subjects, I work with English
a. much less b. less c. almost equally as much d. more e. much more
 23. My English teacher has influenced my English studies
a. very inspiring b. fairly inspiring c. don't know d. negatively e. very negatively
 24. In my opinion we have been taught English
a. very inefficiently b. rather inefficiently c. don't know d. fairly efficiently e. very efficiently
 25. In my opinion my English teacher has been
a. very pleasant b. pleasant c. don't know d. unpleasant e. very unpleasant
Rate how easy/difficult and how pleasant/unpleasant it was to give oral and written answers in the just ended listening comprehension test.
- I found
26. *speaking*: a. very easy b. fairly easy c. don't know d. rather difficult e. very difficult

27. *writing*: a. very easy b. fairly easy c. don't know d. rather difficult
e. very difficult
28. *speaking*: a. very pleasant b. pleasant c. don't know d. unpleasant
e. very unpleasant
29. *writing*: a. very pleasant b. pleasant c. don't know d. unpleasant
e. very unpleasant

THANK YOU VERY MUCH FOR YOUR TROUBLE!

Appendix 4

The schools of the experiment marked on the map of Finland.

Pretesting school

Pirkkalan lukio, Pirkkala (in close vicinity to Tampere)

Schools

1. Kuusamon lukio, Kuusamo
2. Kouvola lyseon lukio, Kouvola
3. Ruotsinsalmen lukio, Kotka
4. Pietarsaaren lukio, Pietarsaari
5. Kirkkopuustikon lukio, Vaasa
6. Svenska samskolan i Tammerfors, Tampere
7. Lammin lukio, Lammi
8. Paimion lukio, Paimio
9. Korson lukio, Vantaa
10. Vehkojan lukio, Hyvinkää



Appendix 5.

A list of observations on the answers per questions. Examples of the answers and scores given in the communicative test.

Question 1.

Ex. They are mainly from 12 to 18 years old.	4
Twelve and eighteen years old.	3
I think they are from 15 to 18.	2
They are 14 years old.	0

Question 2.

Ex. It is a comprehensive school and a Catholic school.	4
It is a comprehension school but Catholic school too.	3
It's a Catholic school.	2
There are many Catholics.	1
The teachers are Catholic.	0

Question 3.

Ex. They must have a letter from their priest to prove it.	4
They must have a letter from priest that their parents are Catholic.	4
They have a letter from priest.	3
They have had a letter from a Catholic priest.	3
They have to be some evidence from priest.	2
They must have a sort of proof which their parents have written down.	1
Their parents must be Catholic.	0

Question 4.

Ex. Because it is not easy to get teachers in some special subjects.	4
Because it is so hard to find Catholic teachers by certain subjects.	3
For example there is no Catholic teachers who can teach German.	2
It is not so many Catholic teachers who have a good education.	1
It's very difficult to teach some subjects.	0

Question 5.

Ex. You don't have to take English but some subject you think it is necessary for your further employment.	4
It's not, it's not necessary if you don't need it in your future education.	4
It is not necessary.	3
It's not, all pupils have not took, take it.	2
It must not be very important.	1
It is very important test.	0

Question 6.

Ex. They have been good in studying French.	4
They have shown some qualification in French.	4
They have been good at French.	4
They've learnt French first.	3
They have learnt French.	2
They are in the second class.	1
They can choose which one of them.	0

Question 7.

Ex. The influence of the family and the Catholic church is stronger than in other schools.	4
Because the school is Catholic and so the parents and also the church have, has better influence to pupils.	3
The influence of family is very strong.	2
They have better discipline because the school is Catholic.	1
Because the religion is very strong at school.	1
They are more disciplined.	0

Question 8.

Ex. In the future there will be no punishments of that kind.	4
They don't use it any more.	3
It is accepted.	2
They don't use it if they don't have to use.	1
There are some police.	

Question 9.

Ex. They have a school council where pupils can influence.	4
They have made a council, a special council.	3
There is a council.	2
They are trying that everybody should have the same things.	0

Question 10.

Ex. They aren't in the near relationship. There's a distance between them.	4
They are not very close.	4
They have not good relationship with teachers and pupils.	3
The pupils is difficult to take contact with teachers.	2
The pupils should talk about their problems.	2
Teacher is higher than pupil.	1
Some people feel that they, that others have more relationship than others.	0

Question 11.

Ex. They are afraid of its high standards.	4
They're are frightened high standard.	3

It has high academic standard.	2
Standard is high.	1
Because it is easier.	0
They don't apply at Oxford university.	0

Question 12.

Ex. They come from grammar schools and public schools.	4
They come from grammar schools.	3
They come from the public schools.	3
They came from public school.	2
From comprehensive schools.	1
They are coming nearby the Oxford.	0

Question 13.

Ex. They can. The comprehensive schools can reach the standard of Oxford.	4
The level is about the same.	4
They are equal good.	3
The academic standard is very high.	2
They are not so high.	0
Oxford's standard is higher than comprehensive schools.	0

Question 14.

Ex. They have to write an essay every week to show what they have learnt.	4
They must write essays every week.	3
They have to write an essay.	2
They have to wrote an essay.	1
They have every week their personal duty.	1
They must to do a little book.	0
They succeed good in examinations.	0

Question 15.

Ex. They have many societies and they have a lot of outside activities.	4
They can belong to various societies.	3
There are many kinds of activities.	2
They can meet very different pupils.	2
They are, have very much opportunities.	1
They learn scientists.	0
They can attend political meetings.	0

Question 16.

Ex. The college system makes them very confined.	4
They are not allowed to go out.	3
Because the studies make them a little confined.	3
The studies take all their times.	2
The college system is so difficult.	2

They are so much to do.	1
It's an old tradition.	0
They are so tired.	0
They are scared to meet too many people.	0

Question 17.

Ex. There are opportunities, but the style is so academic that it is not easy.	4
They were not so good, but it was possible.	4
It's not so good because the course is so academic.	3
It is possible.	2
It is good.	1
They have difficulties to learn Germany.	1
It is easy to learn German.	1
Because all studies happen in German.	0

Question 18.

Ex. He has been in Germany many times, e.g. one year as a language assistant, then as a guest pupil at school.	4
He has visited Germany and he has been a guest pupil in Germany.	3
He had been in Germany for some years.	2
He was in German one year.	2
He had been Germany before.	1
He was in German.	1
He meet guest pupils in school.	0

Question 19.

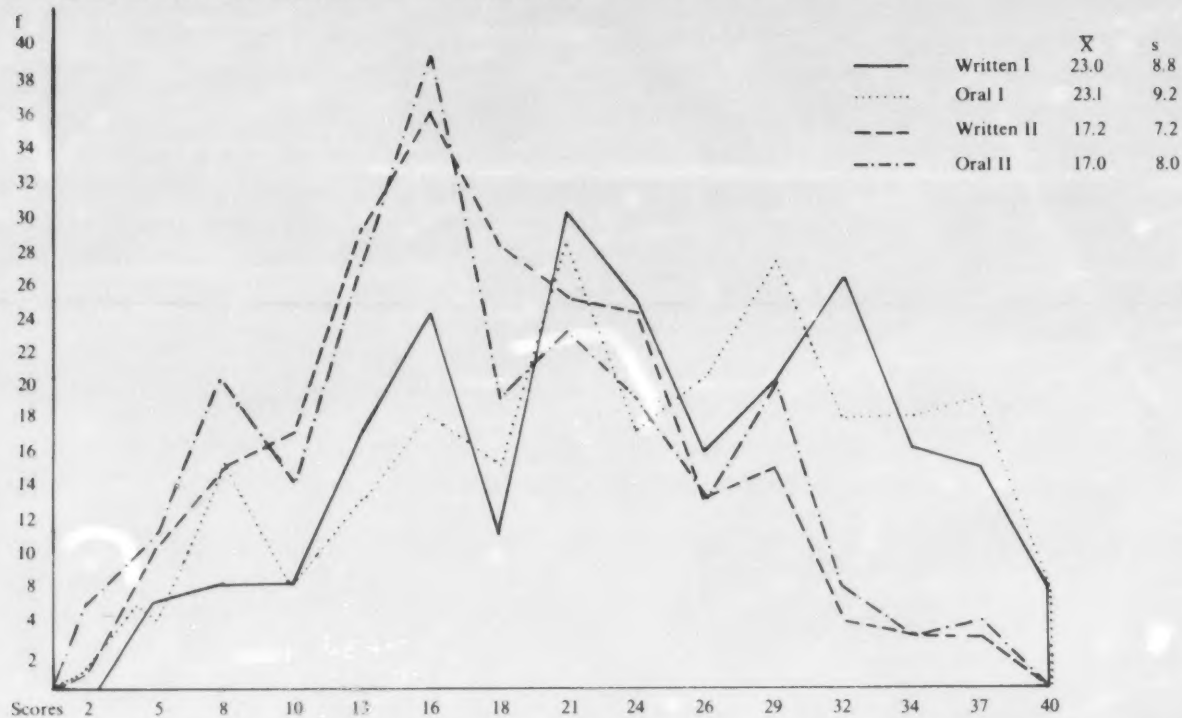
Ex. He's been talking only about one of the Germanies.	4
He has taught them about German meaning only the Federal Republic.	4
He has taught about Germany and only meant the German Federal Republic.	3
He couldn't have told them the truth because he hasn't known that.	2
He hadn't give enough information about the people in German.	1
It's not so democratic as he said.	0

Question 20.

Ex. He has met many new strange people and he has learnt many things about Germany.	4
He has relaxed and learnt new things and met people from other countries.	4
Well, he can both relax and learn and meet people.	3
He had met new people and he's got lot to tell.	2
Intensive course. It's very good indeed.	1
He have met new people and learnt Deutsch.	1

Appendix 6.

A diagram of the score dispersion in the communicative test as four subtests



Appendix 7.

Item analysis of the communicative test as four separate subtests (n = 203).

Item	Written response		Oral response		Written response		Oral response	
	$\bar{X}$	s	$\bar{X}$	s	Average facility value (%)	Correlation of item with total score	Average facility value (%)	Correlation of item with total score
1	2.9	1.5	2.6	1.5	71.4	.40	64.0	.43
2	2.6	1.1	2.6	1.0	63.7	.49	63.9	.42
3	1.9	1.7	1.9	1.7	48.2	.59	46.9	.56
4	3.1	1.0	3.0	1.0	78.3	.49	75.0	.52
5	2.1	1.8	2.1	1.7	51.6	.50	51.7	.48
6	1.8	1.5	2.0	1.6	44.6	.40	50.0	.46
7	2.0	1.4	2.4	1.3	50.9	.55	59.0	.61
8	2.5	1.5	2.6	1.5	61.1	.55	64.4	.63
9	1.5	1.6	1.4	1.6	36.3	.51	35.5	.56
10	2.8	1.3	2.7	1.4	69.8	.43	67.7	.59
					Mean	57.6	57.8	.53
11	2.2	1.4	2.1	1.5	55.1	.49	52.0	.52
12	1.6	1.4	1.7	1.4	38.7	.35	43.2	.41
13	0.9	1.5	1.2	1.5	22.7	.37	29.6	.26
14	2.3	1.3	2.2	1.3	58.5	.59	54.3	.58
15	1.4	1.3	1.4	1.9	33.9	.50	33.5	.56
16	1.4	1.1	1.1	1.1	34.6	.23	27.2	.50
17	2.0	0.9	2.0	1.0	51.0	.45	48.5	.61
18	2.0	1.1	2.1	1.2	50.1	.54	53.3	.65
19	1.1	1.4	1.0	1.4	27.5	.46	24.9	.54
20	2.4	1.0	2.3	1.2	28.7	.59	57.9	.66
					Mean	43.1	42.4	.53

Appendix 8.

Intercorrelations of items for the groups separately.

Items	A																			B																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
2	29																			23																		
3	30	46																		35	34																	
4	23	28	44																	38	23	32																
5	24	19	36	37																23	23	33	30															
6	25	34	30	26	23															26	28	34	26	24														
7	25	31	43	33	39	25														29	31	40	41	38	40													
8	29	30	36	33	37	26	37													32	25	42	41	41	30	51												
9	25	34	41	22	34	30	42	32												23	33	37	31	34	33	42	49											
10	20	23	27	31	30	15	30	48	24											30	31	39	40	37	32	41	48	45										
11	24	42	38	21	36	33	28	38	37	31										26	31	40	35	33	29	41	43	34	30									
12	25	25	35	22	20	29	19	32	30	28	30									18	23	22	23	19	36	19	16	24	27	23								
13	04	20	26	12	18	10	21	10	15	14	14	04								02	19	12	14	18	18	15	25	30	15	28	18							
14	19	44	45	22	24	28	28	37	39	30	44	38	13							28	33	41	46	32	43	48	49	39	42	42	24	25						
15	21	38	48	19	25	29	32	23	41	30	36	33	24	35						20	24	26	33	34	33	35	32	32	30	28	18	28	40					
16	16	32	30	06	26	14	22	17	28	23	23	22	25	41	31					04	14	11	18	15	18	18	16	24	21	17	13	19	17	21				
17	27	38	48	27	34	23	33	25	41	31	42	28	26	41	44	40				18	30	28	29	23	30	40	36	35	38	30	18	23	31	34	15			
18	28	46	47	24	39	31	39	27	45	23	42	35	21	49	42	45	50			28	29	41	39	24	30	36	44	32	32	36	27	25	51	34	10	41		
19	16	31	39	15	19	29	25	22	42	21	36	34	22	34	42	37	40	37		11	24	30	24	24	36	36	32	44	31	28	32	24	36	39	11	26	29	
20	25	47	40	26	29	28	40	28	43	34	45	29	21	48	49	44	49	61	40	32	30	37	42	28	32	37	39	45	40	40	30	27	51	38	24	38	44	29

Appendix 9.

A factor analysis for items of the communicative test.

Principal axes solution						
Item	Written			Oral		
	I	II	III	I	II	III
1	.45	-.10	.13	.49	-.38	.05
2	.56	-.30	.22	.47	-.01	-.28
3	.67	-.16	-.08	.62	-.09	-.11
4	.57	.09	-.11	.58	-.22	.18
5	.56	.16	-.23	.55	.12	.17
6	.46	-.23	.10	.52	-.04	-.29
7	.61	.01	-.20	.68	.05	-.01
8	.63	.25	.16	.70	.11	.18
9	.57	-.15	-.11	.63	.25	-.05
10	.52	.41	.19	.66	.07	.06
Eigenvalue	3.18	.47	.26	3.51	.29	.28
Percentage variance	31.8	4.7	2.6	35.1	2.9	2.8
11	.57	.04	.07	.59	-.20	-.05
12	.41	-.23	-.20	.47	-.20	-.10
13	.42	-.15	.15	.30	.35	-.15
14	.69	.12	-.05	.65	-.18	.15
15	.57	-.12	.05	.52	.03	-.25
16	.27	-.16	.34	.57	.23	.15
17	.54	.15	.05	.67	.10	-.05
18	.65	.25	-.15	.74	.03	.21
19	.52	-.28	-.22	.58	-.00	-.24
20	.68	.09	.11	.75	.06	.14
Eigenvalue	3.0	.31	.27	3.69	.38	.27
Percentage variance	30.0	3.1	2.7	36.9	3.8	2.7

Significance levels of factor loadings using the Burt-Banks formula: I .18, II .19, III .20.

The formula:

$$\text{Standard error of a loading} = \frac{\text{standard error of a correlation}}{\left(\sqrt{\frac{n}{n+1-r}} \right)},$$

where n = the number of variables in the analysis, that is the position of the factor during extraction (Child 1970, 97).

Varimax solution								
Item	Written				Oral			
	I	II	III	h^2	I	II	III	h^2
1	.40	.21	.17	.23	.15	.56	.22	.38
2	.64	.15	.16	.45	.19	.15	.49	.30
3	.49	.17	.46	.48	.33	.33	.43	.40
4	.24	.29	.45	.34	.36	.50	.16	.41
5	.14	.29	.54	.39	.52	.21	.16	.34
6	.48	.10	.19	.27	.21	.20	.52	.36
7	.29	.22	.54	.42	.50	.28	.37	.46
8	.28	.57	.29	.48	.63	.30	.23	.53
9	.40	.12	.43	.36	.56	.08	.39	.46
10	.13	.65	.20	.48	.53	.27	.31	.44
Eigenvalue	1.44	1.07	1.39	3.90	1.84	1.02	1.23	4.09
Percentage variance	14.4	10.7	13.9	39.0	18.4	10.2	12.3	40.9
11	.46	.23	.27	.34	.52	.14	.32	.39
12	.18	.47	.09	.26	.56	.00	.19	.35
13	.22	.23	.35	.22	.02	.47	.13	.24
14	.61	.30	.18	.49	.46	.07	.51	.48
15	.36	.34	.32	.35	.48	.42	.22	.45
16	.08	.05	.45	.21	.14	.34	.51	.40
17	.51	.15	.19	.31	.38	.39	.41	.46
18	.67	.24	.03	.51	.35	.23	.64	.59
19	.23	.57	.14	.40	.47	.37	.20	.40
20	.57	.23	.32	.48	.37	.30	.60	.58
Eigenvalue	1.88	.99	.70	3.56	1.65	.96	1.72	4.33
Percentage variance	18.8	9.9	7.0	35.6	16.5	9.6	17.2	43.3

Appendix 10.

A factor analysis for the questionnaire.

Variable	Principal axes model — 6 factors					
	Factors					
	I	II	III	IV	V	VI
1	.37	-.16	-.32	.15	-.03	.14
2	.30	-.24	-.37	-.32	.19	.15
3	.04	.09	.22	-.06	.05	-.22
4	.10	.15	-.07	.07	.02	.11
5	.15	.24	-.04	.07	-.05	-.15
6	.32	.43	-.27	-.09	.08	-.11
7	.30	.08	-.20	-.01	.22	-.19
8	.22	.38	-.07	.08	.27	-.21
9	.12	.33	-.17	-.06	.13	.12
10	.21	-.21	-.33	-.27	.25	.06
11	.02	-.07	.16	-.19	.18	-.14
12	.32	-.20	.19	-.08	-.02	-.27
13	.23	-.27	.36	-.18	.08	-.26
14	.19	-.27	.15	-.09	.15	-.13
15	.51	.34	-.13	-.02	-.23	-.18
16	.27	.57	.14	-.28	.02	.16
17	.21	.10	-.25	.23	-.19	-.18
18	.43	-.24	-.04	.42	.09	.06
19	.14	.21	.32	.13	.29	.24
20	.11	-.27	-.11	.15	-.06	-.11
21	.58	-.04	.08	.11	.05	-.07
22	.24	-.28	-.12	-.06	.03	-.01
23	.56	-.36	.21	-.14	-.17	.14
24	.52	-.33	.08	-.15	.06	.05
25	.55	-.35	.09	-.01	-.26	.21
26	.30	.47	.17	-.21	-.14	.17
27	.26	.36	.28	.21	.16	.15
28	.38	.42	-.03	-.13	-.28	-.09
29	.43	.06	.09	.42	.13	.03
Eigenvalue	3.10	2.44	1.17	1.00	.73	.70
Percentage variance	10.7	8.4	4.0	3.4	2.5	2.4

Varimax solution — 6 factors

Variable	I	II	III	IV	V	VI	h^2
1	.32	-.17	.27	.17	.10	-.25	.31
2	.23	.02	.62	-.01	-.09	-.03	.44
3	-.05	.07	-.12	.09	.03	.29	.11
4	.01	.08	.04	.10	.13	-.14	.06
5	-.03	.06	-.06	.32	.07	.00	.11
6	-.08	.23	.24	.50	.09	-.06	.38
7	.00	-.06	.30	.31	.13	.11	.22
8	-.20	.09	.12	.40	.29	.11	.32
9	-.12	.25	.19	.19	.14	-.15	.19
10	.11	-.03	.57	-.02	-.07	.03	.34
11	-.03	.05	.08	-.07	-.00	.32	.12
12	.30	-.10	.01	.13	-.01	.37	.26
13	.27	-.04	-.01	-.05	.00	.54	.37
14	.20	-.11	.12	-.07	.05	.32	.18
15	.22	.19	.02	.62	.02	-.02	.47
16	-.02	.65	.03	.26	.15	.02	.52
17	.08	-.18	-.04	.42	-.01	-.16	.24
18	.38	-.34	.08	.11	.40	-.08	.43
19	-.00	.24	-.05	-.10	.49	.06	.32
20	.17	-.32	.04	.05	-.03	-.01	.14
21	.40	-.03	.10	.29	.29	.15	.36
22	.27	-.15	.23	.01	-.04	.05	.15
23	.72	.06	.09	-.03	.04	.16	.55
24	.53	-.01	.27	.00	.09	.22	.41
25	.73	-.01	.06	.01	.04	-.03	.54
26	.13	.58	-.08	.25	.10	-.02	.44
27	.03	.27	-.14	.14	.51	.02	.37
28	.16	.36	-.08	.51	-.05	-.03	.42
29	.23	-.13	-.04	.23	.51	-.03	.38
Eigenvalue	2.32	1.54	1.21	1.86	1.23	0.99	9.14
Percentage variance	8.0	5.3	4.2	6.4	4.2	3.4	31.5

Appendix 11.

Coefficients of congruence, correlation and root mean square between the two factor solutions of the questionnaire for groups A and B separately. (n = 203; decimal points omitted.)

Congruence	Correlation											
	I _A	II _A	III _A	IV _A	V _A	VI _A	I _B	II _B	III _B	IV _B	V _B	VI _B
I _A		-30	22	-01	03	-19	-26	42	-22	39	74	-50
II _A	11		-45	-10	-33	-02	73	-25	53	-32	-40	47
III _A	44	-02		07	-02	-39	-67	58	-01	30	04	11
IV _A	29	24	33		-01	-12	-14	-03	30	48	04	-11
V _A	25	-01	19	21		-15	-19	34	-56	-21	42	-22
VI _A	08	21	-10	12	04		40	-46	-17	24	-14	11
I _B	07	80	-26	15	04	51		-39	26	-14	-27	17
II _B	55	05	67	20	45	-22	-12		-21	13	13	-15
III _B	20	69	31	53	-14	13	47	11		04	-43	-20
IV _B	59	13	51	65	09	41	18	34	39		05	-28
V _B	81	-00	29	30	54	08	03	30	02	33		-44
VI _B	-09	62	34	19	03	28	36	08	45	10	-08	

Root mean square	I _A	II _A	III _A	IV _A	V _A	VI _A	I _B	II _B	III _B	IV _B	V _B
I _A											
II _A	37										
III _A	27	37									
IV _A	28	31	26								
V _A	29	35	27	25							
VI _A	33	32	33	28	28						
I _B	36	18	39	31	32	24					
II _B	24	36	19	28	23	35	37				
III _B	31	21	27	21	32	29	25	31			
IV _B	24	36	24	20	30	25	33	28	27		
V _B	16	38	29	27	22	31	36	29	34	29	
VI _B	34	23	25	25	27	24	26	29	22	30	33